

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 00:18:06 ; Search time 716.17 Seconds
(without alignments)
11842.939 Million cell updates/sec

Title: US-09-269-874a-2
Perfect score: 4940
Sequence: 1 cgcacgcgtatgaatacatc.....ttcatcataatagatcgtatg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_032802:.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4938.4	100.0	4940	19	AAV21451
2	2837.6	57.4	4940	19	AAV35363
3	2706.2	54.8	5760	6	AAAN50530
4	1262	25.5	5181	16	AAO80911
5	1000	20.2	1950	18	AAAT97956
6	989	20.0	1897	18	AAAT97958
7	969.6	19.6	1896	18	AAAT97957
8	744.2	15.1	1065	20	AAAX56008
9	744.2	15.1	1065	20	AAAX5586

10	742	15.0	1140	20	AAAX56021	Merozoite surface
11	742	15.0	1140	20	AAAX5593	Merozoite surface
12	719.8	14.6	1077	22	AAAC68976	Merozoite surface
13	625.2	12.7	1088	20	AAAX56009	Merozoite surface
14	625.2	12.7	1088	20	AAAX5587	Merozoite surface
15	525.2	10.6	1128	22	AAAF9840	DNA encoding major
16	322.8	6.5	1128	22	AAAF9840	DNA encoding major
17	308.8	6.3	1068	18	AAAN81148	DNA sequence encod
18	308.8	6.3	1068	18	AAAN81151	DNA sequence encod
19	246.6	5.0	354	18	AAAT80403	PFMSPI (P19)A codin
20	246.6	5.0	354	18	AAAT94550	PFMSPI (P19)A codin
21	243.4	4.9	737	6	AAAN50355	Storage-specific
22	240.2	4.9	668	9	AAAN82176	31-1 Repeated Dele
23	201.8	4.1	456	9	AAAN81150	DNA sequence encod
24	201.8	4.1	452	9	AAAN81149	DNA sequence encod
25	198.4	4.0	333	22	AAAC68977	Merozoite surface
26	194.8	3.9	387	18	AAAT80404	PFMSPI (P19)S codin
27	194.8	3.9	387	18	AAAT94549	PFMSPI (P19)S codin
28	194.6	3.9	309	24	AAAD22459	Plasmodium falcipa
29	194.6	3.9	3147	24	AAAD22460	PFMSPI.19-human
30	194.6	3.9	3147	24	AAAD22461	Human C3d3 DNA-P1
31	191.4	3.9	309	24	AAAD22462	Plasmodium falcipa
32	191.4	3.9	3147	24	AAAD22463	Mutant PFMSPI.19-h
33	191.4	3.9	3147	24	AAAD22464	Storage-specific
34	119.2	2.4	306	6	AAAN50354	Chicken leucocytos
35	105.6	2.1	3399	17	AAAT05868	Plasmodium vivax m
36	104	2.1	618	22	AAAS00655	DNA encoding signa
37	99.4	2.0	165	18	AAAT93729	DNA encoding signa
38	99.4	2.0	165	18	AAAT10929	Drosophila melanog
39	85	1.7	5688	23	ABLT6543	Drosophila melanog
40	85	1.7	7588	23	ABLT6542	DNA encoding Leuco
41	81.6	1.7	1686	16	AAO87587	Human immune syste
42	81.6	1.7	7758	24	ABLT33103	Synthetic gene VIV
43	72.4	1.5	5438	22	AAAH7054	Synthetic gene VIV
44	72.4	1.5	6101	22	AAAH7055	Plasmodium falcipa
45	70.6	1.4	3579	21	AAA/0099	Plasmodium falcipa

ALIGNMENTS

RESULT 1	
AAV21451	
ID	AAV21451 standard; DNA: 4940 BP.
XX	
AC	AAV21451;
XX	
DT	23-SEP-1998 (first entry)
XX	
DE	P. falciparum modified gp190 DNA.
XX	
KW	gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW	monoclonal antibody; passive immunisation; parasite; ss.
XX	
OS	Plasmodium falciparum.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	10..4929
FT	/*tag= a
XX	/product= gp190
PN	W09814583-A2.
XX	
PD	09-APR-1998.
XX	
PF	02-OCT-1997; 97WO-EP05441.
XX	
PR	02-OCT-1996; 96DE-4404817.
XX	
PA	(BUJA/) BUJARD H.
XX	
PI	Bujard H, Pan W, Tolle R;

XX WPI: 1998-240088/21.
 DR P-PSDB: AAM54145.
 XX Recombinant production of complete gp190/MSP-1 Plasmodium surface
 PT protein - useful in anti-malaria vaccines, also stabilising genes by
 XX reducing their AT content
 XX
 PS Example 1; Fig 3c; 48bp; German.
 XX
 CC This sequence encodes a modified Plasmodium falciparum gp190/MSP-1
 CC (merozoite surface protein) which has a reduced AT content resulting in
 CC a higher stability of the protein. Such a protein is useful in
 CC vaccines against malaria or for producing monoclonal antibodies (for
 CC passive immunisation). The complete gp190 protein can now be produced
 CC outside the parasite and has, at least over extended regions, the native
 CC pattern of folding. Larger amounts of the protein can be produced
 CC recombinantly than would be possible using the parasites as source.
 CC
 XX Sequence 4940 BP; 1669 A; 1156 C; 1060 G; 1055 T; 0 other;

Query Match 100.0%; Score 4938.4; DB 19; Length 4940;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;
 Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 cgcacgcgtatgaataatcatcttctcctcgttcattctcgtttttatcatcaact 60
 Db 1 cgcacgcgtatgaataatcatcttctcctcgttcattctcgtttttatcatcaact 60
 QY 61 cagtcgctgaaccacgaatccctatcagagcgtgttaagaatctggaagat 120
 Db 61 cagtcgctgaaccacgaatccctatcagagcgtgttaagaatctggaagat 120
 QY 121 gccgtcttaccggaacgcgtgtccagaagaagatgtgtctgaatgaaggag 180
 Db 121 gccgtcttaccggaacgcgtgtccagaagaagatgtgtctgaatgaaggag 180
 QY 121 gccgtcttaccggaacgcgtgtccagaagaagatgtgtctgaatgaaggag 180
 Db 121 gccgtcttaccggaacgcgtgtccagaagaagatgtgtctgaatgaaggag 180
 QY 181 agtgcacgcgcgttcaacaacacacgcgttctaaggctctgtgctagcgtgag 240
 Db 181 agtgcacgcgcgttcaacaacacacgcgttctaaggctctgtgctagcgtgag 240
 QY 241 tccggtgggtctgtgctcctcgtgggttcctgcgtccggcgagcgtgtgcatagt 300
 Db 241 tccggtgggtctgtgctcctcgtgggttcctgcgtccggcgagcgtgtgcatagt 300
 QY 301 ggcctcgtgcgaacgcgtgtccgggaacgtcgaagaaccatcctgtgcaact 360
 Db 301 ggcctcgtgcgaacgcgtgtccgggaacgtcgaagaaccatcctgtgcaact 360
 QY 361 agcgattccgaacgcgtcctcgaacgcgtcgaacgcgtggaatactatctcctc 420
 Db 361 agcgattccgaacgcgtcctcgaacgcgtcgaacgcgtggaatactatctcctc 420
 QY 421 actatcaagaagatgaagacacacgcgtgttgacgtcctatataatgtcgacag 480
 Db 421 actatcaagaagatgaagacacacgcgtgttgacgtcctatataatgtcgacag 480
 QY 481 tctgataacatcattgcgttcaaatatctgattgacgttgcgaagaagatcaatga 540
 Db 481 tctgataacatcattgcgttcaaatatctgattgacgttgcgaagaagatcaatga 540
 QY 541 ctgtacaagtgaattctacttgaactgtgtaaggccaactgaatgaagcttgcgc 600
 Db 541 ctgtacaagtgaattctacttgaactgtgtaaggccaactgaatgaagcttgcgc 600
 QY 601 aatgactatgtcaaatctcattcaattggaagatcagagccaagcttgagcgtatg 660
 Db 601 aatgactatgtcaaatctcattcaattggaagatcagagccaagcttgagcgtatg 660
 QY 661 aagaattgtcttcgatacgcgaagcctctgcgaacaatcaagaacatgtgggaag 720
 Db 661 aagaattgtcttcgatacgcgaagcctctgcgaacaatcaagaacatgtgggaag 720

QY 721 atggaattatataataaagaataagaagaccatcgagaacatacagctgactcga 780
 Db 721 atggaattatataataaagaataagaagaccatcgagaacatacagctgactcga 780
 QY 781 gaatccaaaagaacacatagacaanaataagaatgcacaagaggaagaagaag 840
 Db 781 gaatccaaaagaacacatagacaanaataagaatgcacaagaggaagaagaag 840
 QY 841 ttgtaccaggccagacgacctgtccatctataaacaacgcttaagaagccataac 900
 Db 841 ttgtaccaggccagacgacctgtccatctataaacaacgcttaagaagccataac 900
 QY 901 ctatcagagctatgtagaagcgcatagacacccctcaagaagaatgaataatcaaga 960
 Db 901 ctatcagagctatgtagaagcgcatagacacccctcaagaagaatgaataatcaaga 960
 QY 961 ctgtcgcgaagaatgaatgaatgaagaatcctcgcgcgaacacacacccct 1020
 Db 961 ctgtcgcgaagaatgaatgaatgaagaatcctcgcgcgaacacacacccct 1020
 QY 1021 aacacgctgtgcgaagaagaagaagatagagcagcagagaagaatcaagaagatc 1080
 Db 1021 aacacgctgtgcgaagaagaagaagatagagcagcagagaagaatcaagaagatc 1080
 QY 1081 gccaaaaccattgaatcacaatagatctcctcttactgaccccttgagctgagatc 1140
 Db 1081 gccaaaaccattgaatcacaatagatctcctcttactgaccccttgagctgagatc 1140
 QY 1141 tacttgaagagaagaataagaatataagacatctccgcgaagaatcgagaagaatca 1200
 Db 1141 tacttgaagagaagaataagaatataagacatctccgcgaagaatcgagaagaatca 1200
 QY 1201 accgaaccatataatataatcccaatggtgtagacgtaccctctgctataacgatac 1260
 Db 1201 accgaaccatataatataatcccaatggtgtagacgtaccctctgctataacgatac 1260
 QY 1261 aacgctcacaacgagctcaatagcttcggtgactgtattaacccctcgattacgaa 1320
 Db 1261 aacgctcacaacgagctcaatagcttcggtgactgtattaacccctcgattacgaa 1320
 QY 1321 gaaccccttaagaatataatcccaatggtgtagacgtaccctctgctataacgatac 1380
 Db 1321 gaaccccttaagaatataatcccaatggtgtagacgtaccctctgctataacgatac 1380
 QY 1381 gagaagaatcaaatgtagaagaagaatgtagaagtcgaagaagaatgtacgaagacgc 1440
 Db 1381 gagaagaatcaaatgtagaagaagaatgtagaagtcgaagaagaatgtacgaagacgc 1440
 QY 1441 agcaaaagtcataacgatacctcaagaagatgtagaagaagctgtgcaacgagatct 1500
 Db 1441 agcaaaagtcataacgatacctcaagaagatgtagaagaagctgtgcaacgagatct 1500
 QY 1501 tccaataatcaaatataatcgcgtcgaacacacacacacacacacacacacacacac 1560
 Db 1501 tccaataatcaaatataatcgcgtcgaacacacacacacacacacacacacacacac 1560
 QY 1561 tcttaacaagtggaagaacgtgacacacacacacacacacacacacacacacacacac 1620
 Db 1561 tcttaacaagtggaagaacgtgacacacacacacacacacacacacacacacacacac 1620
 QY 1621 cataactctgagaagctcacaagaagctcttaagtataatgagacatctctcgcgaac 1680
 Db 1621 cataactctgagaagctcacaagaagctcttaagtataatgagacatctctcgcgaac 1680
 QY 1681 attgtgtgagaagaagaataagatataagaagaatcctaagaatgaagtcgaaacgag 1740
 Db 1681 attgtgtgagaagaagaataagatataagaagaatcctaagaatgaagtcgaaacgag 1740
 QY 1741 atcgaaacgtctgttagaacaatgaagaagatgaagaacagctgtgttagaagaagatc 1800
 Db 1741 atcgaaacgtctgttagaacaatgaagaagatgaagaacagctgtgttagaagaagatc 1800

QY 1801 acaaaagcgaataaataaccagatgagaaatccctggaggtcccgatattgttaaaagtc 1860
|||||
Db 1801 acaaaagcgaataaataaccagatgagaaatccctggaggtcccgatattgttaaaagtc 1860
QY 1861 caagtgcagaaggtgtcctccatgaaacagaatgtgaactcaagaagactcaactcatt 1920
|||||
Db 1861 caagtgcagaaggtgtcctccatgaaacagaatgtgaactcaagaagactcaactcatt 1920
QY 1921 ctgaaagaagcttgaggttaaaataataatatactgtccgaaatagttaaaagagaat 1980
|||||
Db 1921 ctgaaagaagcttgaggttaaaataataatatactgtccgaaatagttaaaagagaat 1980
QY 1981 aagcagaagaacataactacatccatcgtactcaagaagaagatagacaactgaaggtcttc 2040
|||||
Db 1981 aagcagaagaacataactacatccatcgtactcaagaagaagatagacaactgaaggtcttc 2040
QY 2041 atgcccgaagctcgagagccttgatcaacgaagagaagaacatlaaaactgaagagcaag 2100
|||||
Db 2041 atgcccgaagctcgagagccttgatcaacgaagagaagaacatlaaaactgaagagcaag 2100
QY 2101 tcaagataactccgaagccttcacagaagagaataaccggaagcgttacccaagaagccc 2160
|||||
Db 2101 tcaagataactccgaagccttcacagaagagaataaccggaagcgttacccaagaagccc 2160
QY 2161 ggaacaagaagccggttcagctctcgaagagcagatagcgtgcaagctcaagcaagaagcaag 2220
|||||
Db 2161 ggaacaagaagccggttcagctctcgaagagcagatagcgtgcaagctcaagcaagaagcaag 2220
QY 2221 aagcagaagcagagcctccaggtgcaggtgcaggttccagaagcgttaagctcaaggtcctaca 2280
|||||
Db 2221 aagcagaagcagagcctccaggtgcaggtgcaggttccagaagcgttaagctcaaggtcctaca 2280
QY 2281 ccacgaagcctctgtgaaataaacaagaacgagaatgtccagcaaacgtgactacattgaaag 2340
|||||
Db 2281 ccacgaagcctctgtgaaataaacaagaacgagaatgtccagcaaacgtgactacattgaaag 2340
QY 2341 ctctatagcttctctgaaataatccatcacatctgcacacaataatataccctcgtctccagc 2400
|||||
Db 2341 ctctatagcttctctgaaataatccatcacatctgcacacaataatataccctcgtctccagc 2400
QY 2401 actatgaaagagaagaatctttaaacaagtacaagaataacaagaagagagtaaaactg 2460
|||||
Db 2401 actatgaaagagaagaatctttaaacaagtacaagaataacaagaagagagtaaaactg 2460
QY 2461 tccctctgtatccacgtgcgacctgtgttcaaatctccagaacaacattcccgattatgtat 2520
|||||
Db 2461 tccctctgtatccacgtgcgacctgtgttcaaatctccagaacaacattcccgattatgtat 2520
QY 2521 tctatgtctgatatagcctcaacaattctctctcaactgttcatgagatatatgagaag 2580
|||||
Db 2521 tctatgtctgatatagcctcaacaattctctctcaactgttcatgagatatatgagaag 2580
QY 2581 gagatgtctgcgaacctgtataaaactccaagaacagacaagaattaaagaacctcttgag 2640
|||||
Db 2581 gagatgtctgcgaacctgtataaaactccaagaacagacaagaattaaagaacctcttgag 2640
QY 2641 gaagcttaagaaggtgtccacactcgtttaaactctctccagctccatgacaacacctg 2700
|||||
Db 2641 gaagcttaagaaggtgtccacactcgtttaaactctctccagctccatgacaacacctg 2700
QY 2701 tctctcacactcagaagaagccggaagtgagcgttaagcagacacactcctcatcagcc 2760
|||||
Db 2701 tctctcacactcagaagaagccggaagtgagcgttaagcagacacactcctcatcagcc 2760
QY 2761 aaccttaataactcaactgaaactgtttgagaacaatccctgtctccggaagaataagaaac 2820
|||||
Db 2761 aaccttaataactcaactgaaactgtttgagaacaatccctgtctccggaagaataagaaac 2820
QY 2821 atctacaagaagaacttatgagcagaagaatcgtccgagaactcttaacgagaagataactgaaa 2880
|||||
Db 2821 atctacaagaagaacttatgagcagaagaatcgtccgagaactcttaacgagaagataactgaaa 2880
QY 2881 gacagcgacacatctataacagagagcttcaactcgtgaataataagccgatgat 2940
|||||

Db 2881 gacagcgacacatctataacgagagagcttaactacttcgtgaatatctaaagccgatgat 2940
|||||
QY 2941 atcaactctcttaacgatgaaactcaaacgttaagaagcttgaaagagacataaagactg 3000
|||||
Db 2941 atcaactctcttaacgatgaaactcaaacgttaagaagcttgaaagagacataaagactg 3000
QY 3001 aagaagaacactgcgaactgcgaagcttgacactgtataacaagaactcaaacgtgaagaga 3060
|||||
Db 3001 aagaagaacactgcgaactgcgaagcttgacactgtataacaagaactcaaacgtgaagaga 3060
QY 3061 ctctctcgaagaagaagacagctcgcgaagtataagatgacagatcaagaagcttgactctg 3120
|||||
Db 3061 ctctctcgaagaagaagacagctcgcgaagtataagatgacagatcaagaagcttgactctg 3120
QY 3121 ctcaagaagcagcttgaaagaagaactcaactcactgtgaacaactccgaaacagatctcag 3180
|||||
Db 3121 ctcaagaagcagcttgaaagaagaactcaactcactgtgaacaactccgaaacagatctcag 3180
QY 3181 aactctcaggtgtcttcaacaagaagaagagccgagatcgcgcggaagacagaacact 3240
|||||
Db 3181 aactctcaggtgtcttcaacaagaagaagagccgagatcgcgcggaagacagaacact 3240
QY 3241 ctggaagaacaccaaagatctctctcaacaacatacaaaagcctcgtcaagatataatagtc 3300
|||||
Db 3241 ctggaagaacaccaaagatctctctcaacaacatacaaaagcctcgtcaagatataatagtc 3300
QY 3301 gagtctctcctctgaaagactctctccgagaagagatccagaccgagataaactagccc 3360
|||||
Db 3301 gagtctctcctctgaaagactctctccgagaagagatccagaccgagataaactagccc 3360
QY 3361 agcctcgagaacttcaagagctctgtctaagctgaaagcgaagctgaaagagaacactagac 3420
|||||
Db 3361 agcctcgagaacttcaagagctctgtctaagctgaaagcgaagctgaaagagaacactagac 3420
QY 3421 ctggaagaagaagaagctcagcttaactctctctagcagactgcataccactgcgcgagctc 3480
|||||
Db 3421 ctggaagaagaagaagctcagcttaactctctctctagcagactgcataccactgcgcgagctc 3480
QY 3481 aaggaagctcatlaagaagaagaactataccagcgcaatagcccaagcgaagaataatacagac 3540
|||||
Db 3481 aaggaagctcatlaagaagaagaactataccagcgcaatagcccaagcgaagaataatacagac 3540
QY 3541 gtgaataacgacactgcgaactcttaagaagaagcttccgtctgaaagagaacagatgtccgact 3600
|||||
Db 3541 gtgaataacgacactgcgaactcttaagaagaagcttccgtctgaaagagaacagatgtccgact 3600
QY 3601 gtgtgtctgaaatctgtgtccgacaacactggaagcagcttcaaacctaaagacactgcatct 3660
|||||
Db 3601 gtgtgtctgaaatctgtgtccgacaacactggaagcagcttcaaacctaaagacactgcatct 3660
QY 3661 actcaatgtcgaagcagactccaatacaaatcaacatctcagaacactgcagatgtgagtc 3720
|||||
Db 3661 actcaatgtcgaagcagactccaatacaaatcaacatctcagaacactgcagatgtgagtc 3720
QY 3721 gataagcttcatcatgtgcacatcttcggaagagcgaagagagactcagatgactcggc 3780
|||||
Db 3721 gataagcttcatcatgtgcacatcttcggaagagcgaagagagactcagatgactcggc 3780
QY 3781 caagtggtcacaacggttgagagctgtcatcctctcgttgtataacatctgttccaacatc 3840
|||||
Db 3781 caagtggtcacaacggttgagagctgtcatcctctcgttgtataacatctgttccaacatc 3840
QY 3841 gagaagaataacgaagtgctctatactgaaacactctggaagcgtctataggtctctcaag 3900
|||||
Db 3841 gagaagaataacgaagtgctctatactgaaacactctggaagcgtctataggtctctcaag 3900
QY 3901 aaacagcttgagaagaataacgttatgacttcaatgtaacagtgaaagagacattcttaacagc 3960
|||||
Db 3901 aaacagcttgagaagaataacgttatgacttcaatgtaacagtgaaagagacattcttaacagc 3960
QY 3961 cgcttlaaagaagaagaatcttcaagaagcgtcttgagaagcgaactgattccctataaa 4020
|||||

Dh 301 ggtcagttgttcagggtgttcaggttaattcaagacgtacaacgttcagataattca 360
Qy 361 aggcattccgcagcgaagttcctacgagcactcaagcgcaggtgagaaactatctcct 420
Dh 361 agtgaattcaagatgcaaaccttaacgctgagatttaaaacacagatgaaatactctgtca 420
Qy 421 actatcaagagcgctgaagttacccaagttgttcgacccttaactatcatgtcgacatg 480
Dh 421 actatcaagagcactcaaatatccctaacctcttgatcttaactaatcatatgttaacttg 480
Qy 481 tgtgataacattcatgtctcaaatatctgattgagcgggttcgaaagatcaatgaactc 540
Dh 481 tgtgataacattcatgtctcaaatatctgattgagatgagataagaaatcaattaa 540
Qy 541 ctgtaacaagttgaatttctacttacttgcactgttgaaggccaacgtgaagatgttcgccc 600
Dh 541 ttatataaatttaaacttttatttctgtatttaagaagcaaaatttaaatgagtgaatgtc 600
Qy 601 aatgactatgtcaaatctcaattccatttgaagatcagagccaagatgtgacgtatg 660
Dh 601 aatgactatgtcaaatctcaattccatttgaagatcagagccaagatgtgacgtatg 660
Qy 661 aagaagtgtgtctcgagatcgcgaagcctctcgacaacatcgaagagacatgtgaggaaag 720
Dh 661 aaaaactgtgttcgagatagagaacacattagaacattataagataltgtaga 720
Qy 721 atggaagaatttatataaagaataagaagccatcgcgaacatttaagacgtgatcga 780
Dh 721 atggaagaatttatataaagaataagaagccatcgcgaacatttaagataltgtaga 780
Qy 781 gaattccaaaagacatagacaanaaataagaatgcaaacgaagagagaaagaagaag 840
Dh 781 gaagatgaagaacacatattgataaataataagaatgcaaacgaagagagaaagaagaag 840
Qy 841 ttgtaaccagggcccgatgacgctgtccatcatacaaacgcttgaagaagccataac 900
Dh 841 ttatacaagcagctcaatagatcttcttctatctacataaacaattagaagaagcatalat 900
Qy 901 ctctcagcgtactcggaagcgcatagacacccctcaagaagaatgaataatatacaaga 960
Dh 901 ttaataagcgcttttagaataaacgcatatgacacctttaaaaaaaatgaaacatttaagaa 960
Qy 961 ctgctcgacaagatlaaataaataagaatcctccgccaagcacaactctggaacacccct 1020
Dh 961 ttacttgaataagatlaaataaataagaatcctccgccaagcacaactctggaataaccca 1020
Qy 1021 aacacgctgtctgcaagaagacaagaatagagagacgagaaagagatcaagaagatc 1080
Dh 1021 aatactctccttgaataagaacaaaaaaltcgaggaaacgacgaataaagaat 1080
Qy 1081 gccaaaaccatlaagttaacataagatctctcttacttgaagtcacctgagctgagatc 1140
Dh 1081 gccaaaaccatlaaattlaaataatgatagttattacttaccacttgaattgaataac 1140
Qy 1141 tacttggagagaaagataaagataatagacatctccgccaagtcgaaagtcgaagaagatca 1200
Dh 1141 tatttaagagagaaaaaataaataatattgataaagtgcgaaggttgaacacaagaatca 1200
Qy 1201 acggaacctaataatcccaatgtgtgtgacgtacccctgtcttataacgatatcaac 1260
Dh 1201 actgaacccaatgatatcccaatgaggttactatccctgttcataatacagatatattac 1260
Qy 1261 aagctctcaacgagctcaaatagcttcgtgtacttgaattaaaccccttcgattataagaa 1320
Dh 1261 aatgtcttaaatgaaacttaattcttcttgggtatlaaattcaattgattatacaaaa 1320
Qy 1321 gaaccccttaagaatctatacagacagacatagagaaagagttatcaagaagaatcag 1380
Dh 1321 gaaccaagtaaaaaacatatatactgataatgaaagaaaaaattcatataatgaaattag 1380
Qy 1381 gagaagatcaaaatttgagaagaagaataatvgagagtgacaagaagaagttacgaagacgc 1440
Dh 1381 gaaaaaatttaaatagaaaaaataaattgaaatctgataaaaaattcttacgaagacaga 1440

Qy 1441 agcaaaagttcaacgatatcaactaaagatgataaagctgtcgaaacgatatcatgat 1500
Dh 1441 tctaagcttataaataatgataatacaaaaagataatgaaataacttaataatgatat 1500
Qy 1501 tccaattcaacaataaataatgcagctgcagcacaactctgcaaaaatgattgagaaacggtac 1560
Dh 1501 agcaaatccaataaataatagatttaacttaatttcgaaaaaataatgattgagaaagatat 1560
Qy 1561 tcttaacaagtggagaaactctgacacacataatattcttgcatactcctatgagaattctaa 1620
Dh 1561 tcatataaagttgagaaactcttaacacacataacttcttgcatactcctatgagaattctaa 1620
Qy 1621 cataacttgaagccttaccacccaagcttgaagataatgagagactattctctcgaggaac 1680
Dh 1621 cataacttgaagccttaccacccaagcttgaagataatgagagactattcttctgaaggaat 1680
Qy 1681 attgttggagaaagaactaaagatattacaagaatcctcataagtaagatcgaaaacgag 1740
Dh 1681 atagatgttgaagaaagaatttaaatatataataaatttaataagcaaaatagaaatgag 1740
Qy 1741 atcgagacgctgttggagaaacttaagaagatggaagacagttgtttgagaagaagatt 1800
Dh 1741 atggaacattagttgaaataatataaagaatggaagacagcttcttgaaaaaaataat 1800
Qy 1801 acaaaagcgcgaataataaacccagatgagagatccctgaggtctccgatatgttaagtc 1860
Dh 1801 actaaagcgcgaataataaacccagatgaaataatttaagaagatctcgacatgtlaaagta 1860
Qy 1861 caagtgcagaaggtgtctcctcatgacacaagaattgataagcacaagaagactcaactat 1920
Dh 1861 caagtgcagaaggtgttatttaattgaacaaaattgscgagatttaaaaagaacccaattgt 1920
Qy 1921 ctgagaagcgtgtgagtttaaacataatatacatgtgcggaatagttaagcaggaagat 1980
Dh 1921 ttaaaaaatgtagaatttaaaaacataatatacatgttcccaattcttaacaaagaagaat 1980
Qy 1981 aagcagagacacatctacatcctatcgtaactcaagaagaagatagaaacctaagaatgttc 2040
Dh 1981 aagcagagacacatctatattatattatgtgttgaagaaaaaagattgataaatttaagaatgt 2040
Qy 2041 atgccaagaatcgagagccttgatcaacgagaagaagaagaacatlaaacgtgaagacag 2100
Dh 2041 atgccaagaatcgagagccttgatcaacgagaagaagaagaacatlaaacgtgaagacag 2100
Qy 2101 tcaagatactccgagccttccacagaaagagagataacccggaacggtccacccaagccc 2160
Dh 2101 tcaagatactccgagccttccacagaaagagagataacccggaacggtccacccaagccc 2160
Qy 2161 ggaacacagcgcttcaagctctcgaaagcgatagcggtgcgaagctcaagcacaagaagag 2220
Dh 2161 ggaacacacagcgcttcaagctctcgaaagcgatagcggtgcgaagctcaagcacaagaagag 2220
Qy 2221 aagcagcagcagcctccagltgcagltgcagltccagaaggtlaaagttcaagltgctlaa 2280
Dh 2221 aagcagcagcagcctccagltgcagltgcagltccagaaggtlaaagttcaagltgctlaa 2280
Qy 2281 ccaccagctcctgtgataatacaagaacgagaatgtcagcaaacctggaactccttgagaag 2340
Dh 2281 ccaccagcagcaggttaataataaactgaaatgttcccaaatagattatccttgaanaa 2340
Qy 2341 ctctatggtctccgtaagatcctacatctgcacacaatatatcctcgtctccacagc 2400
Dh 2341 ctctatggtctccgtaagatcctacatctgcacacaatatatcctcgtctccacagc 2400
Qy 2401 actatgaagagaagatcttaataacagatacaagaataaacaagaagagagagatnaactg 2460
Dh 2401 actatgaagagaagatcttaataacagatacaagaataaataaagaagagagagatnaactg 2460
Qy 2461 tccctgtgtgataccagtggaactcgtgtcaatatccagaacaacatcccggtatgtat 2520
Dh 2461 agtcaatgtgataccatltagacttatgttataatacaaaaataacatacctgtatgtat 2520

Db	3601	gltctgaagctgaagctgacccgaacacattagaacaaagctcaaccgaagaaacccagcatca	3600
Qy	3661	actctatcgcgagccgagctcccaatatacattacacactctcaagaacgtcgacgatgagctc	3720
Db	3661	actctatcgaggaagcagagctctaacacataacacactctcaaaaagctcgatgataagta	3720
Qy	3721	gatgacgtacatcattgtgtcctactcttcgcgcgagcgcgagagactacgaatgacctgcgc	3780
Db	3721	gatgacgtacatcattgactactactctttgagaatccgagaagatattatgatattga	3780
Qy	3781	cagcgcgcacccgcttgagcgtctgcactccctccgcgttgattgataacattctgcacaaac	3840
Db	3781	caagtaagtaacaggaagacgcagtaacccctccgtaattgataacactctctcaaaac	3840
Qy	3841	gagaacgcatacgcagagctctactctgnaaacctctgcgcgcgtctctataggtctccag	3900
Db	3841	gaaatctgatatgaggtttattatatttaaacaccttgacgcggtgttatagaattaa	3900
Qy	3901	aaacagctcgagaaatacgtatgataccttcaatgtcacaacgtgaaggacattctgaacgc	3960
Db	3901	aaacattatgaaataacagttatgacatttaagttatgataagatatttaattca	3960
Qy	3961	cgctttaagagagaanaatttccagaagcgtcttgagagcgacttgattccctataa	4020
Db	3961	cgatttaataacgtgugaanaattccaaaagttcttgaaatcagatcttcaatccataaa	4020
Qy	4021	gacctgacctctcaactcagttctgtctaaagacccaataaagttccatataaagagag	4080
Db	4021	gatttaacatcagaattatgtgtgcacaagatccatatataattctcttaataagaanaa	4080
Qy	4081	aggagataattctgtctagtttaacatacataaagagctccatgcgacccgatatcaat	4140
Db	4081	agaagataattcttaagcagtttaattatataaagattcaatgatatacgatataaat	4140
Qy	4141	ttcgctaagtgtgcgcggggtatataagaatccctgagcgaanaatacagcgtcact	4200
Db	4141	ttcgaaaatgtcgtctctgtgatatataaatactatccgaaaataataacagattta	4200
Qy	4201	gactctataaagaatataccaacgaataagcagagcagaaatgaaaataatcgtccctc	4260
Db	4201	gattcaattataaanaataataccaacgaaaacagaggaatgagaataactccctc	4260
Qy	4261	ctgaataacatcgaaacacctgtataaagacgtgaacgcanaaatcgcacctctcgtaatt	4320
Db	4261	ttaacaacattctgacacttataataaacgattaagataaattgattatttgaatt	4320
Qy	4321	caacctgagcgcaaggtccctcaacttacttaagaaagaaagcaatggtgaaattaaatc	4380
Db	4321	catttaagaagcaaaagttcttaaatatacaatacgaatacaacgctagaattaaata	4380
Qy	4381	aagagcgcgaactaccccaaaacatccaagaacagctgcagattccagaanaaatatc	4440
Db	4381	aaagacacttaattacttaaaaaacatccaagacaataatgtgcgatttttaaaaaaatatc	4440
Qy	4441	aatttcgtcggaattgcagacctgtctacggatataacacacaacatctcgtccag	4500
Db	4441	aatttcgttggaattgtctgatttatacaacgattataacacataaactattgacaag	4500
Qy	4501	ttctgtccactgcagtggttccgnaaacctcgcnaaacgltgctgagcaattgtctc	4560
Db	4501	ttccttataagataatggttttggaaaactctgtcaaaacggttatactaatattact	4560
Qy	4561	gacgcgaacccgcagggcagctgctbaaacatctccagcaccaaatggtgataaagacgac	4620
Db	4561	gatgaaaactctgcagatgattataaacttccacacacccaatgctataaaaaaatatgtc	4620
Qy	4621	ccccgaataagcgcgttttcagcatctgnaacgacgcggaagctgaaatgtctcctgt	4680
Db	4621	ccacaacattctgtatgttccagacatttagatgaaaggaagaaatgttaaatgtttata	4680
Qy	4681	aactacaacaagaagataagttgcgtgagaaaccaaaccttacctgatagaaac	4740

OY	1090	atcaagttcaacatagatctctctcttaactgatacccttgagctyggatctaacttgaga	1149
Db	1296	atlaaatltaatltagatagttatttaactgatacccatgtaattagaataactttaaga	1355
OY	1150	gagagaataagaatlagagcattccgcgaagttcgagagcaagaagaataccgaac	1209
Db	1336	gaaaaaaataaatacttgataaagtgcgaagtttgaaaaaagaaatcaatgnaacc	1415
OY	1210	aataatatacccaatggtgagacgtacccctgtctcttaaaagatatcaaacgcttc	1269
Db	1416	aatgatatccaatgtgagttacttaacctgttgatataacgatatataacaatgctta	1475
OY	1270	aacgagctcaatlagctcggltgactgtaaacccctcgatatacgaagaacctct	1329
Db	1476	aatgacttaattctttgttgatttaattcaatccattgtaattatacaaaagaccagt	1535
OY	1330	aagantcttaaccagcaatlgagagaagaagtttatcaacgaataccaagagaagtc	1389
Db	1536	aaaacatatacttgataatgaaagaaaaaattcaataatgaaatlaagaaaaatt	1595
OY	1390	aaaattgagaagaagaataattgagtgatgtaacaagaataatgaaagccgcgaag	1449
Db	1596	aaaattgaaaaaataaaatctgtaactctgtataaaaaacttaagaagacagatctaag	1655
OY	1450	ctaaacatactacataaagatatagaaagctgtctgaacgagatcatgatccaattc	1509
Db	1656	ttaattgataatacaaaaagataatgaaaaattcaattgaattatgatagcaattc	1715
OY	1510	aacaaataactcgacctgcaccaacttcgcgaaaaatgatatgaggaacggtactctta	1569
Db	1716	aataataatagattttaacttaatttcgaaaaaaatgatatggtgataaagatatcata	1775
OY	1570	gtggagaacactgacaacacataactctgtgacacctgaagaaattccaagaaatct	1629
Db	1776	gttgagaacttaacacaccttaacttcttgtaacctgaagaataatccaacaatactt	1835
OY	1630	gagaagctcaccaagactcttaagatataatgagagactatctctgcggaacaattgtg	1689
Db	1836	gaaaagattaaacaaagctcttaataatataatgagaagattactctttaagaaata	1895
OY	1690	gagaagaagactaaagattttcaagaatctcataagtaagaatcgaaaacagagtcga	1749
Db	1896	gaaaagagattaaaataattttataaatacttaataagaaaaaataagaataatgagattg	1955
OY	1750	ctttctgagaacactlaagaagatgtaagaacagttctgtgaaagaaagattacaaga	1809
Db	1956	ttagtgtgaataataattataaaaaagatagaacagcttttgataaaaaaatactata	2015
OY	1810	gaaaataaacccagatgagaagactcctcgaggtctccgatatgtttaagttccaag	1869
Db	2016	gaaaataaacccagatgagaagaaattttagaagatctgcgactgtttaaaagttcca	2075
OY	1870	aagttgctcccatgaaacaagatttgatgtaactcaagaagaactccaactctgaaga	1929
Db	2076	aaagttcttaataatgaaacaataatgtgcgaattataaaaagaccatgattttaaaat	2135
OY	1930	gttgagatttaaaaaataatacatalatggtgcgaatagttataagcagagataagcaga	1989
Db	2136	gtagaattaaacactaataataactgtgtcccaattcttcaaaaaaagaataatagaag	2195
OY	1990	ccatactactccatcgtactcaagaagaagatagacaacatgaaagtgtcaatgccaa	2049
Db	2196	ccattattatttaattgtgttgaaaaaagaagattgataaattaaaagtgtcaactga	2255
OY	2050	gtcgagagcctgatacaacgcgaagagaagaataatataaacctgaagagagatcgata	2109
Db	2256	gtagaatcatgtataaatatgaaagaaaaaataaataaacaacaaggtcaactcgata	2315
OY	2110	tcgagagccttccagaagaagagatataaacgcgcgacaggtctacacccaagcccgga	2169
Db	2316	tcggaacatacaaacgcgaagagaataataacgcgaagcaactacaacacctggaata	2375
OY	2170	gcgcggttcaggtctcgaaagcgatagctgtgcaagttccaagcaagaagcagaagcga	2229

Db	2376	gcaagatctgtctttagaaggagatctcagtaacaagcacaagacaagaacaaaacaagca	24335
Qy	2230	caagccttcagtcgcaagtcgcccgtctccagagcgccaagctcaagttgctcaccaccagct	22885
Db	2436	caaccacccagtaaccagtaaccagctcccgaaagcaaaagcacaaagctcccaacaccacagca	24995
Qy	2230	cctgtgaaataacaagaccggaatgctcagcaaacctgacactctgagagctctatgag	23495
Db	2436	ccagtaaatataataaactgtgaaatggttcccaaatagattatctctgaaaaattatcaaa	25555
Qy	2350	ttcctgaaataaccctacatctctgcacaataatatcctcgtctctacagcactatgac	24005
Db	2556	ttttaaatcttcaatataatagtccacaataatatttggtttcacacactcaactatgac	26155
Qy	2410	gagaagatctctaaacgctcagaatagatagacaaagagagagatgaatctgcccctgt	24665
Db	2616	gaaaagatataaacaacatataaatatacaagaggagaaagaaacaaatlaagttcagct	26755
Qy	2470	gaccacatgagaccgctgttcacatcaccagacaacacattccgcttatgtatctatgttc	25255
Db	2676	gatcatattagacttatgtttatatacaaaataacatacactcgttaatgtatctatgttt	27335
Qy	2530	gatagcctcaaca-----atctc	25475
Db	2736	gatacgttaaacacatagttatcacaacatatattggtttatgtaaaagaanaattggttat	27955
Qy	2548	ctctctcaacgcttcacagagatgatatgagaagagatgctgcgcaactgatataaactc	26075
Db	2796	gattatataacttaactaggaatataagaaaagaatgggttgttaattatataaactc	28555
Qy	2608	aaagacaacgacaagattaaagacactctgagaggaagctaaagaaagctccacactctgt	26675
Db	2856	aaggttaatgcaaaattaaaatattatagaggagcgaaaaagtatccacactctgta	29155
Qy	2668	aaactctctcttcagctccacatgcaaccacgtctctcacacctcaagacaagcccgaa	27225
Db	2916	aaactctcttcaagttaatcaatgtcaaacacttatcatlaaacactcagatagaataaccgaa	29755
Qy	2728	gtgagcgttaacgacgacactccacatctgacccaacttaataactcactgaactggttt	27875
Db	2976	gtaagtgaatgtatgatacatcaccatctcacaactttgaataatagtttaaatatttt	30335
Qy	2788	gagaacacccgctctcctcgcgaagaataagacaatccacccaagaacttatgtgacagaa	28475
Db	3036	gaaacacatattgagctctgtgaaaaaaacaaaatataaccagaagtaatagtgtaaaaa	30955
Qy	2848	tcgtcccggaactctctacgagaaagatctctgaaaagacagcgacacatctcataagagcg	29075
Db	3096	agtagtgaaaaactttatgaaaagatataaaagatagtgatcatcttataatgaactc	31555
Qy	2908	ttcacttaacttcgtaaatctcaaaagccgagatcagatccaactcttaagatgaactcaa	29675
Db	3156	tttacaatatttgttaaaaatctaaagcctgagatataatcatctgatatgatatcaaaa	32155
Qy	2968	cgtaagagcctgagagaggaatcaataaagctgaaagagacactgtgaactgtgacctgac	30275
Db	3216	aggaagaatttagaagagatataatataaattaaaaaaactttacagtatcatcttgat	32755
Qy	3028	ctgtacaacaagtcaaaaactgtgaactgtgagagactcttcgacaagaagaagcagctgcg	30875
Db	3276	ttataataataataatataaattgaagaagtatttgtataaaaagaaaacagttggt	33355
Qy	3088	aagtaataagtcgagatcaagaagtgtgacctgtctaaaggagaggttgaagaacaaactc	31475
Db	3336	aaataataaagcaatttaaaaaacttaactttataaagaaacaaattgaaatacaaatg	33955
Qy	3148	aactcaacgacaacatccgaaacacgtatctcgaaactctccagttgctcttcaacaagaag	32075
Db	3396	aattcaacttaataacccaagaagcagtataccaacactttctgttttctcttaacaaaaa	34555
Qy	3208	aaggaagccgagatccgcgagacgagaaacactctgagaaacacaagaattctctctcaaa	32675

D	3456	aagaaagctgaataatagcaagaactgaaacacattagaanaacacaaaataattattgaa	3515
Q	3268	cactacaagaagcctcgtcaaglatataatggcgagctctctcctcgtgaagctctcc	3327
D	3516	cattataagagcttgtaataattataatggttgatcatctccattaaaacttaagt	3575
Q	3328	gaggagaagcatcagaccgagataactcggcagcctcagactctaaaggtccgtct	3387
D	3576	gaagaatcaaatcaaacagagaataattctgcagtttagaanaactttaagattaa	3635
Q	3388	aagctcgaaagcgaagcgtgaaggaacccggaaccccgagagaagaagctgaactc	3447
D	3656	aaactgaagaagaaataataaaggataactcnaacttgaagaagaaaaaatcatcacta	3695
Q	3448	ctctgcgagactgcataccactgcgcgcgcgcgaaggaagtgatcataagaacagactac	3507
D	3696	tcaagaggtttaactcttaactttaattgcttgatataaagaagataataaaaatacaaat	3755
Q	3508	accggcaatgcccaagcgaataataacagacgtgataacgcacatgcacatgttaaca	3567
D	3756	acagatattctccaaagcgtgaataatacgaatgtaaacaaatgcacattagaactctaca	3815
Q	3568	aagctcctgcctaaagaanaagatgcgcgcacatggtgtgtcgaactctgcctccgaca	3627
D	3816	aaattctccgaagaagaacagatgttgtaacaacagttgtaagtgtaagtggaatccgaca	3875
Q	3628	ctggagcagctcaaccataaggaagcctctgatactcatgtctcgagccgaattccataca	3687
D	3876	ttagaacaagaatccaaacaaagaacacagatccaactcatgtagcgagctctaacaca	3935
Q	3688	attaccacatctcaagaactcgacgatgtgcgatgacgtcatatattgtgctatcttc	3747
D	3936	ataacaacatcacataaaatgctcgatgtgaagatagatgcagatcatcatatatt	3995
Q	3748	ggcgaagcgaggagagactacgaatgcacccgcgcgcaggtggttaacggcgagcgctca	3807
D	3996	ggagaaatccgaagaagatcttgatgttttaggaacaagtacgtacagcggaagcagta	4055
Q	3808	ccttcgcgtatgtataacattctgtccaaatcgcgaagaacatacgaagtctcatctg	3867
D	4056	acttcgtaattgatacaactcttctaataattgaaatgtaatgaggtttattat	4115
Q	3868	aaacctcgtgcagcgctctataagctctcacaagaaacagctggagataacgtatgac	3927
D	4116	aaacctttagcaggtgtttatagaagttttaaanaaaacatagaanaataacgtatagca	4175
Q	3928	ttcaatcgaagtgaagaagacattctgcgaacacgcgtttaataagaggaataattca	3987
D	4176	tttaatgttaatgtltaagatattttaaatltaacagatttaataaacytgaataattca	4235
Q	3988	aagcctctggagcgagctgtgattcctcataaagccagacctccctcaactacgtgtc	4047
D	4236	atgtttttaaatcagatttaactatcatataaagatttaacatacgaatataatgtgttc	4255
Q	4048	aaggaaccatacaagttcctcaataaagaagaagagataaattctgtctagtataaac	4107
D	4296	aaagatcatataaattctttaataaagaaaagaagataaattcctaagcattataat	4355
Q	4108	tataaagaagcccatctgacacacgtatcaatlttcgataagatgtgtggtattac	4167
D	4356	tataitaagatccaatagataagatataaattctgcaaaatgactgtcttgatattat	4415
Q	4168	aagatccctggcgaaataatacaagctctgcaccttgactcatataaagtatatacaagat	4227
D	4416	aaaaattatccgaaaaataataatcagatttagatttcaattaaaaataatacaagac	4475
Q	4228	aagcaagcgagaaatgaaaaatactgcctctcctcgtgaatcaaatcgaaactgcgttaca	4287
D	4476	aaacaagtgtaaatgagaataactctcccttttaacaataattgtgaccttatataa	4535
Q	4288	acagtgaaagacaataatcgacctcttcgtcaatcaactcggagcgcaaggtccatacat	4347
D	4536	acagtttaatatataatttgattattctgaattcatltagaagcaaaagttcttaaat	4595

OY	4438	acttaacgagaagagcaatctggygaagttaataatcaagagctgaactactctcaaacatc	44407
OY	4438		
Db	4596	acatatggaatctcaacatcagtagaagtttaataataaagaacttaacttaaaaaaatc	46555
OY	4408	caagacaagctgcygagattccaagaanaataacaatttcgtcggaattgcagactgct	44675
Db	4656	caagacaatctggagattcttaaaaaaaatacaaatcttcgttcggattgcgtattatca	47735
OY	4468	accgattatacccaacaacatctcttcagacccaagtcttcgtccactgycatggtgtcga	45227
Db	4716	acagattataaccataataactattatgcaacaagttctccttagtaacagratgtgtttgaa	47757
OY	4528	aacctgcaccaaaagatgctgcgcaatctgctcga-cggcaacctgcgaagcgatcgtga	45866
Db	4776	aactgtctaaatctcgctttatcttaatttacttgattggaacttcgcaagratgtttaa	48359
OY	4587	catctccagagacaacatgctgttaagaagaacagt--cccccagaatagcggctgttcag	46444
Db	4836	catctcaacaacacaaatgctgtataaaaaaaacaatgatccaacagagttcgtgatgttcaa	48955
OY	4645	catctgcgcgagcgcggaagctgcaagtgtctctcgtgaactacaacaagaagagataag	47044
Db	4896	catctagatgtagaagaagaagatgtaaatgtttatctaataacacaagaagtagtaaa	49555
OY	4705	tgctgtgaggaaccccaaaccttactctgcaatgaagaacaaatgggggtgtgacgcgatagt	47644
Db	4956	tgtgtgtgaatctcaatctcaactctgtgaaacgaataaatacgtgtgagatgcagatgcc	50155
OY	4765	aaatgcacgcggaagaagacgaacggtctctaaacgagaagaagaatacacatgcaggtactaag	48244
Db	5016	aaatgtacgcgaagaagattcagtagcaacgcgaagaagaataacatcaatcgtaagtactaaa	50757
OY	4825	ccgactctctacacctcttcgcagcgggatttttgcctcaagcttcaatttccttggtatc	48844
Db	5076	ctgattgtttctccactcttcgatgtgaattttctgcagttcctctcaactcttaagtgaata	51355
OY	4885	tcctctgcgcgcgaacccatgcagctgcgatacctgtacagcttcacatctaatagatgatg	4939
Db	5136	tcattctattataactccatcagttcaataataataacagtttcatttaaaaaatgtagg	5190
RESULT 4			
ID	AAQ80911	standard; cDNA; 5181 bp.	
XX	AAQ80911;		
XX	24-AUG-1995 (first entry)		
XX	DE	Plasmodium falciparum MSA-1 gene cDNA.	
XX	KW	Plasmodium falciparum MSA-1 gene; recombinant poxvirus;	
XX	KW	multicomponent multistage malarial vaccines; immunogens;	
XX	KW	malaria diagnosis; ss.	
OS		Plasmodium falciparum (p486195).	
XX	PN	W09428930-A.	
XX	PD	22-DEC-1994.	
XX	PF	10-JUN-1994; 94WO-US06652.	
XX	PR	11-JUN-1993; 93US-0075783.	
XX	PR	09-JUN-1994; 94US-0257073.	
PA		(VIRO-) VIROGENETICS CORP.	
XX	PI	De Taisne C, Paoletti E, Tine JA;	
XX	DR	WPI: 1995-036113/05.	

OY	3814	gtgtttataaacattctgtcccaaaatcggagacgaatcgaatgtctctatctgaaact	3873
Db	4009	acaattgataacttcctctcgaagtttgaaatgataagatgtatataattlaaaact	4068
OY	3874	ctgcagacgctctataagctctctcagaagaaacgcctgagaaataacgttgatgactccat	3933
Db	4069	ttagcttggatgatatagaagctttaaaaaaacaattgaaaaaacaatttttaacttcaat	4128
OY	3934	gtcaacgtgaaagacattctgaaacgcgctttaaagaagagaaattccaagaacgtc	3993
Db	4129	ttaaatttgaacgatatctttaaattltaacgtctttaaagaaacgaaaatttcttagatgta	4188
OY	3994	ttgagagcgacttgatttccctctataagccgcgactcctcaactagttgtgcaagac	4053
Db	4189	ttgaatctgattttaaigcaatttlaaacatatacctccaagaatgaataattatggaagt	4248
OY	4054	ccatacaagttccctcaataaagagagagggataattctgtctcagttacaactatc	4113
Db	4249	tcatttaattatgtgaattcgaacaaaaaacacacttttaaaagtttaacaatatata	4308
OY	4114	aaggaatccatcgacacgcgaatalcaatttcgtcgaatgagtctggtgtattacaagatc	4173
Db	4309	aaagaatcagtagaanaatgataattaaatttgcacgaagagatagattatgaaag	4368
OY	4174	ctgagcgaaaaatacacagctctgccttgctctatataaagttatatacagaataag---	4230
Db	4369	gttttaagcaataataaaggtatgtattgagatacaatttaaaaagttatcaagaagaaag	4428
OY	4231	-----ca	4233
Db	4429	gagagttcccatcatcatcaaccaacaacaactcgtcaacggtataaaacgaagaaaca	4488
OY	4234	ggcgagaaatgaaaaatactctgcctctcgtgaataacatcgaaacccgttacaagacgtg	4293
Db	4489	aagaagagaaagtaagttccttcctcatatttlaacaacaactgtgacacttatacaataactta	4548
OY	4294	aagacaaacaatcgacccctctgcgaatttcgcggcggaaggtcctaactatactaac	4353
Db	4549	gttaataaaatgacgactcttaacttaacttaagggcaagaatlaacagattgttaatggt	4608
OY	4354	gagaaagacaaatctgagaggttaaaatacgaagcgctgaatacctcaaaacaatccaagac	4413
Db	4609	gaaaaagatgaaagacatgtttaaataacttaacttaacttagtattaaagcaattgtgac	4668
OY	4414	aagcttgcagattccaagaaaataacaatttcgtcgaattgcgaccgtgtctaacgat	4473
Db	4669	aaaaatagatcttlttlaaaaaacataacgactcogaagcaattlaaaaaaatgatabaatgat	4728
OY	4474	tataacacaacaatctcctcgaccgaagttctgtccactgcgacgtgttctgaaaaactc	4533
Db	4729	gatacgaanaaaagatatagctctggtgcaaatattacttagtacggagttag---ttcaaatattt	4785
OY	4534	ggcaaaacagttcgtgagcaatctgcctcgacggcaacgtccaaggtacgtgtaacatccc	4593
Db	4786	ccataatacaataatacaaaataattgaagagaaaaattccaagataatgttaaacatttca	4845
OY	4594	cagacacaatctgctgaaagaaacagttgcccccgaagatagcgtgtttcaagcactgtgac	4653
Db	4846	caacacacaatgtcgtaaaaaaacaatgtccagaanaattcctgagtttccgaacatttagat	4905
OY	4654	gagcgcgaagatgtcaaggtgtctctcgtgaactacaacaagaagagataaagtgcgtgag	4713
Db	4906	gaaagagaaatgttaaatgtttatttaaatlttaicaacaagaaggtgtataaagtgttgaa	4965
OY	4714	aaccaaacctacctgcgaatgnaaaaacatgtgcggtgtgagcgcgagtgtcaaatgcaac	4773
Db	4966	aatccaatcctactgtlaacgaaaataatgtgtgagctgtgaagcagatgccaaatgtacc	5025
OY	4774	gaggaagacacgcggtcttaacggaagaataacaatgtgaggtgtactcaagcccgcaatcc	4833
Db	5026	gaagaaagatctcaggttagcgaacggaagaaataacatactgtgaatgtactcaaacctgattct	5055
OY	4834	tatcacactcttcgacgggaatttttgcctcagctcaaatlcttcgtggcatcctcttcctg	4893

Db	5086	tatccacttttcgatgtatatttcttcgcagttctctcaacttctttaggaatattccta	5145
Qy	4894	ctgaatccatgcgcacccgtctacacgttcatcatca	4929
Db	5146	ttaatcatcatgtaataattataacagtttcatttaa	5181
RESULT	5		
AAAT97956			
ID	AAAT97956	standard; DNA; 1950 BP.	
AC	AAAT97956;		
XX			
DT	03-APR-1998	(first entry)	
XX			
DE	Chimeric MSA-1 antigenic protein 1 used in a malaria vaccine.		
XX			
KW	Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;		
KW	MSA-1; recombinant vaccinia virus; Plasmodium falciparum; anchor peptide;		
XX	immune response; humoral; cell-mediated; merozoite; ss.		
XX			
OS	Chimeric - Mammalia.		
OS	Chimeric - Plasmodium falciparum.		
OS	Chimeric - Homo sapiens.		
XX			
Key	Location/Qualifiers		
FT	1..165		
FT	/*tag- a		
FT	/note- "signal sequence added to improve		
FT	immunogenicity"		
FT	misc_feature	166..1893	
FT	/*tag- b		
FT	/note- "Plasmodium falciparum MSA-1 peptide"		
FT	misc_feature	1894..1950	
FT	/*tag- c		
FT	/note- "anchor sequence added to improve		
FT	immunogenicity"		
XX			
PN	W09726911-A1.		
PD	31-JUL-1997.		
XX			
PE	29-JAN-1997;	97WO-US01395	
XX			
PR	29-JAN-1996;	96US-00593006	
PA	(GEQU) UNIV GEORGETOWN.		
XX			
PI	Davidson EA, Yang S;		
XX			
DR	WPI; 1997-393372/36.		
XX			
PT	Malaria vaccine - comprises expression vector expressing fragment of		
PT	merozoite surface antigen		
XX			
PS	Claim 33; Fig 2; 75pp; English.		
XX			
CC	The present sequence encodes a chimeric protein that contains a		
CC	mammalian signal and anchor sequence, and a merozoite surface		
CC	antigen-1 (MSA-1) carboxy terminal peptide. The C-terminal fragment of		
CC	MSA-1 provides a more specific response than the complete MSA-1,		
CC	and attachment of anchor and signal sequences improve the immunogenicity		
CC	of the protein better than the use of an adjuvant. The chimeric protein,		
CC	and expression vectors (analogues that express MSA-1 without either		
CC	signal or anchor peptides), particularly in the form of recombinant		
CC	vaccinia virus, are used in vaccines to prevent or treat malaria caused		
CC	by Plasmodium falciparum. The vaccinia vector expresses the antigen		
CC	fragment for many days, or even years, generating a long-lasting immune		
CC	response (humoral and/or cell-mediated) against the merozoite form of		
CC	the parasite, in humans or other animals.		
XX			
Sequence	1950 BP; 830 A; 236 C; 277 G; 607 T; 0 other;		

Db 999 aaattcaaaaatgttttgaatcagatttaattccataataaagattacatcaagtaa 1058
 QY 4038 ctacgtgtgcaggaccatatacagttccctcaataaagaagaggataattctgtc 4097
 Db 1059 ttatgtgtcacaagatccatataatttcttaataaagaagaaagataattcttaag 1118
 QY 4098 tagttacaactatcaatcagactccatcgacacgcgataatcctcgtatgtgtc 4157
 Db 1119 cacttaataattatgaagattcgaatagatcgataataaatttgcgaatgtctct 1178
 QY 4158 ggggtatatacagatctcgaagcaaaatacagaatctgcctcgtatcatttaaaagta 4217
 Db 1179 tggatattataaattatccgaaaaataataatcagaattagattcaataaaaaata 1238
 QY 4218 tatcaacgataagcagcgagataaataataatctgcctcctcctcaataacgcgaac 4277
 Db 1239 tatcaacgataaagcgagaaatgagaataatcctcctcctttaaacaatactgaac 1298
 QY 4278 cctgtacaaagacagtgacgacaaaatcgacctctcgtatcctacgtgagcgcaagt 4337
 Db 1299 cctataataacagttatgataaattgtatttcttgcattcattagagcaagat 1358
 QY 4338 cctcaactactactaagaaagacgaatgtggaagttaaatcaagagactgaactact 4397
 Db 1359 tctaaattatcatatgagaaatcaaacgtgagaagttaaaataaagaacttaattact 1418
 QY 4398 caaaacaatccaaagaaagctgcgacattccaagaaataaacaattcgtcgaaatgc 4457
 Db 1419 aaaaacaatccaaagaaatgtgcgacattccaagaaataaacaattcgtcgaaatgc 1478
 QY 4458 agaacctgtctacagatataacacaacactcctcgtacaaagtctcgtcgaatgc 4517
 Db 1479 tgattatatacagattataacacataataactatgacaaagtctcgtacaggtat 1538
 QY 4518 ggtgttcgaaaacctgcgaacaaacagtgctgagacactcgtcgcgacacactgcagg 4577
 Db 1539 ggtttttgaaaatctgtcgaacaaacgcttattactaatttactgtgagaaactgcagg 1598
 QY 4578 catgtgaacatctccacgacccaatgcgtgaaagaacagtgccccaagatagcgctg 4637
 Db 1599 tatgttaaacatttcacacacccaatgcgttaaaaaacaatgcccacaatctcgtatg 1658
 QY 4638 ttccagatctcggagacgcggaagatgtaagtgtcctcctgaactacaaagaagag 4697
 Db 1659 ttccagacactttagatgaaagaaagaaatgtaatttcttaataatcaacaagaag 1718
 QY 4698 agataagtcgtggaagacccaacctcctcgaatggaataaacaatggcggtgagc 4757
 Db 1719 tgataaagtgttgaaatacccaatccctactgttaacgaaaaataatggtgagatgtgc 1778
 QY 4758 cgatgtcaaatgcaacggaagaaagacagcgcttaacggaagaagaaatcacatgcagtg 4817
 Db 1779 agatgccaatgtaacggaagaagatcaggtagcaggaagaagaatcacatgtaagt 1838
 QY 4818 tactaagccgactcctcctacactcctcgaagatcttctgcctcagcttaatt 4873
 Db 1839 tactaagcgtattcctacacacttctcgtatgtatttctgcagttcctcctaact 1894

OS Synthetic.
 XX PN WO9920774-A2.
 XX 29-APR-1999.
 XX 20-OCT-1998; 98WO-US22226.
 XX 15-MAY-1998; 98US-0085649.
 XX 20-OCT-1997; 97US-0062592.
 XX (GENZ) GENZYME TRANSGENICS CORP.
 XX Chen LH, Meade H;
 XX WPI: 1999-288313/24.
 XX P-PSDB: AA09372.
 PT Modified malarial protein for use in anti-malarial vaccines
 PS Claim 1; Fig 1; 35pp; English.
 CC This novel, modified nucleic acid encodes the 42 kDa C-terminal
 CC portion (see AA09372) of malarial merozoite surface protein MSP-1
 CC (MSP-1-42), an important target for the development of a vaccine
 CC against Plasmodium falciparum. The nucleic acid sequence has been
 CC modified compared to the native sequence of MSP-1-42 (see AAX56009)
 CC such that 306 nucleotide positions have been replaced to lower the
 CC AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability
 CC motifs while maintaining the same protein amino acid sequence.
 CC These alterations allow MSP-1-42 to be expressed in mammalian cell
 CC culture and in transgenic mice. Native MSP-1-12 is known to be
 CC difficult to express in cell culture systems, mammalian cell
 CC culture systems or in transgenic animals. The invention allows
 CC expression of MSP-1-42 protein in the milk of transgenic animals,
 CC and also provides a DNA vaccine comprising a vector containing the
 CC altered MSP-1-42 sequence.
 XX Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;
 SX
 Query Match 15.1%; Score 744.2; DB 20; Length 1065;
 Best Local Similarity 81.3%; Pred. No. 7.1e-171;
 Matches 863; Conservative 0; Mismatches 198; Indels 0; Gaps 0;
 QY 3799 gctgtacactcctcgtatgataacattcgtcccaaatcgagacgaatacgaatg 3858
 Db 1 ggcgtacactcctcgtatgataacattcgtcccaaatcgagacgaatacgaatg 60
 QY 3859 ctctatcgaacactcgtgcagcgctctataggtctctcaagaacacgtgagaataac 3918
 Db 61 ctgtacctgaagcgcgtgcaggggtctaccgagcctgaagaagagctgagagaacac 120
 QY 3919 gtagatccttaatgtcaagtcagtgaggaattcgaagacgcgtttaaagaagaga 3978
 Db 121 gtagatccttaatgtcaagtcagtgaggaattcgaagacgcgtttaaagaagaga 180
 QY 3979 aattcaagaacgcttctgagagcgactgattccctcaataaagacgcctccttaac 4038
 Db 181 aactcaagaacgcttctgagagcgactgattccctcaataaagacgcctccttaac 240
 QY 4039 tacgtgtcaagaacccaatacagaatctcctcaataaagaagaggaataattctgtct 4098
 Db 241 tacgtgtcaagaacccaatacagaatctcctcaataaagaagaggaataattctgtct 300
 QY 4099 agttacaactatatacagaacccatcgacacccagataatctcgtatgtgtc 4158
 Db 301 agttacaactatatacagaacccatcgacacccagataatctcgtatgtgtc 360
 QY 4159 ggggtattacaagatcctgcgcaaaaataaagatcgtgacctgaacttaaaagat 4218
 Db 361 ggggtattacaagatcctgcgcaaaaataaagatcgtgacctgaacttaaaagat 420

RESULT 8
 AAX56008
 ID AAX56008 standard; cDNA; 1065 BP.
 AC AAX56008;
 XX
 XX 31-AUG-1999 (first entry)
 DE Merozoite surface protein MSP-1-42 modified cDNA.
 XX
 XX MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; codon usage; transgenic animal; ss.
 XX
 OS Plasmodium falciparum.

Db	481	ctgtacaagaacgctcaacgctaagattgattctgttcttgatcatcacactggaagccaaagtc	540		
Qy	4339	ctcaactatctactaacgagaagagcaatctgtggaaagttataaatacaagagctgtcaactctc	4398		
Db	541	cttgactaccacatactgtgaaagagcaacgtttgtgagttcaaatgatcaagagctgtgatttctt	600		
Qy	4399	aaaacatcccaagaacagctgcgcagattcttaagaanaataatacaattctgtcgaaattgca	4458		
Db	601	aagcgcattccagagataagctgcgcgatttccaagaagaatacaaacacttcgttcggatgcgc	660		
Qy	4459	gaactgtctaacgattataacccaacaactctctgacccaagtctctgtlcaactgtgacatg	4518		
Db	661	gattctgagcacccgattacaacccaacaacactgtctgaccaagtctcctgagcacccggtatg	720		
Qy	4519	gtgtctgaaaaacccctgcgaacacagctgtgaggaatctgtctgacgcgcgaactgtcaagcc	4578		
Db	721	gtctctgaaaaacactgtgcgaacacccgttccctgagcaacctgtgtgatacgtgaacctgtca	780		
Qy	4579	atgctgtgacaactctccacgacccaactgtcgttgaaagaaacagltgcgcccaagatagcggctct	4638		
Db	781	atgctgtgacaactcagccagccacgctgtgtgtaagaagcagltgtccccaagacagcgggtct	840		
Qy	4639	ttcgggactctggaacgcgcgcgaagagtgtaagtgctctcttaactatcaacaagaaga	4698		
Db	841	ttcgcgacaacccgtgtgagagagagaggtgtcaagtgctgtcgtgaactacaagacgaagatc	900		
Qy	4699	gataagctgcgttggaagaaccccaaacacttaccctgcaatgaaacaatagtcgggtgtgaccc	4758		
Db	901	gataagctgtgtggaagaaccccaactctactctgttaacgagagacaatgtgtgatatgtcc	960		
Qy	4759	gattgtctaattgcgaacgcgcggaggaagacagcgcgtctcaacgcgaagaagaataatcatagc	4818		
Db	961	gattgtctaattgcgaacgcgcggaggaatltcaagtgaggaacaaggaatgtacactgtgagtc	1020		
Qy	4819	actaagccgcgactctcatccactcttcgcagcgattcttg	4859		
Db	1021	accagcctgattctctatccactcgttcgatgtatctctg	1061		
RESULT 10					
AAK56021	ID AAK56021 standard; cDNA; 1140 BP.				
XX	AAK56021:				
XX	31-AUG-1999	(first entry)			
DE	Merozoite surface protein MSP-1-42 modified cDNA.				
KM	MSP-1; merozoite surface protein; malaria; vaccine.				
KM	protein engineering; protein expression; codon usage;				
KM	transgenic animal; mutant; ss.				
XX	Plasmodium falciparum.				
OS	Synthetic.				
XX	Key	Location/Qualifiers			
FT	CDS	1..1131			
FT	sig_peptide	/*tag- a			
FT	mat_peptide	1..45			
FT	mat_peptide	/*tag- b			
FT	mat_peptide	46..1128			
FT	mat_peptide	/*tag- c			
XX	W09920774-A2.				
XX	29-APR-1999.				
XX	20-OCT-1998;				
XX	98MO-US22226.				
XX	15-MAY-1998;				
XX	98US-0085649.				
XX	20-OCT-1997;				
XX	97US-0062592.				

PA	(GENZ) GENZYME TRANSGENICS CORP.
XX	
PI	Chen LH, Meade H:
XX	
DR	WPI: 1999-28833/24.
XX	
DR	P-PSDB: AAY09374.
XX	
PT	Modified malarial protein for use in anti-malarial vaccines
PS	
XX	
PS	Example; Fig 11: 35pp: English.
XX	
CC	This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC	part (see AAY09374) of malarial merozoite surface protein MSP-1
CC	(MSP-1-42), an important target for the development of a vaccine
CC	against Plasmodium falciparum. The nucleic acid sequence has been
CC	modified compared to the native sequence of MSP-1-42 (see AAX5009)
CC	such that the AT content has been reduced and 10 mRNA instability
CC	motifs eliminated while maintaining the protein amino acid sequence.
CC	In addition, a sequence encoding a 15-amino acid beta-casein signal
CC	peptide has been added to the 5' end of the sequence, and N262Q
CC	and N181Q mutations have been introduced to eliminate
CC	N-glycosylation sites. These alterations allow MSP-1-42 to be
CC	expressed in the mammary gland (i.e. milk) of transgenic mice. The
CC	invention also provides a DNA vaccine comprising a vector containing
CC	an altered MSP-1-42 sequence.
CC	
XX	
SO	Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other:
Query Match	15.0%; Score 742; DB 20; Length 1140;
Best Local Similarity	80.8%; Pred. No. 2.5e-170;
Matches 865; Conservative	0; Mismatches 205; Indels 0; Gaps
OY	3791 ccggtgagcgcttaccctccgttgatgatacaattctgcacaaatcgagacgaat 3850
DB	38 ccatitgagccgttaccctccgttcgcatcgataacatcctgtccagatcgagaacgagt 97
OY	3851 acgaagtcctctatctaaacccctctgcagcgctctatagctctcacaagaacacgtcg 3910
DB	98 acgagtgctgtacctgaagccctctgcagagatctcaacggcgcttgaaagcgcgtg 157
OY	3911 agaaatacgtgatgaccttcaatgltcaacgtgtaagacatctctgaacagcgccttaata 3970
DB	158 agaacaacgtgatgaccttcaacgtgtaagacatctctgaacagcgcgttcaaca 217
OY	3971 agaaagaaatttcaagaagctcttgtagagacgacttgttccctataaagccctgact 4030
DB	218 agagggagaaacttcaagaagctcttgtagagacgacttgttccctataaagccctgacta 277
OY	4031 ccttaactacgtgttcaagaagccacatacagttctctcataaagaagaagagataat 4090
DB	278 gcagcaactacgtgttcaagaagcctccatacagttctctcataaagaagaagagataat 337
OY	4091 ttctgtctagtatacactatataatcaaggactcgcacacgcatatcaatttcgtcaatg 4150
DB	338 tctcgagcagcttatacactatataatcaaggactcgcacacgcatatcaatttcgtcaatg 397
OY	4151 atgtgtctgtggttatcaagaatctctgaacgaaataatacaagcttcaacttgactctata 4210
DB	368 atgtctctgtggttatcaagaatctctgaacgaaataatacaagcttcaacttgactctata 457
OY	4211 aaaagtatatacaagaataagcgaagcgagatgtaaaataatactgtccctctcgtgaataca 4270
DB	458 agaagtatatacaagaataagcgaagcgagatgtaaaataatactgtccctctcgtgaataca 517
OY	4271 tcgaacctgtatacagaacgtgaaacgacaaatcgacctcttgtaattcaactcgtgag 4330
DB	518 tctgagacctgtatacagaacgtgaaacgacaaatcgacctcttgtaattcaactcgtgag 577
OY	4331 ccaaggtctcctaactatacttaagagaagaagcagatgttggaagttaaatcaagggcgtga 4390
DB	578 ccaaggtctcctaactatacttaagagaagaagcagatgttggaagttaaatcaagggcgtga 637

Db 781 atgtgaacattgcccaaccaatgcgttaagaagcaatgtccacaacatccgagatgt 840
 Oy 4639 ttcaagcactctgagcagcgacgaagtgcaagtctctccgaactatacaacaaga 4698
 Db 841 ttcaacacatctgagcaggaagaatgtaagtctgtctgaactacaagaagaagt 900
 Oy 4699 gataagtcgtgaggaacccaacctactctgaactgaagaatgaggtgtgaccc 4758
 Db 901 gataagtcgtgaggaacccaacctactctgaactgaagaatgaggtgtgaccc 960
 Oy 4739 gatgtcaaatgcacgcaggaagacagcgctctaaacggaagaatcaacatgcagatgt 4818
 Db 961 gacgtcaagtgcacgcaggaagacatctgtcttaacggaagaatctactgtcgagatgt 1020
 Oy 4819 actaagccgcagctctatccactcttgacggagattttgtctcagctcaatt 4873
 Db 1021 actaagccagactctacccttctgtcgtgaactcttctctcctaact 1075

RESULT 13

AAK56009
 ID AAK56009 standard; cDNA; 1088 BP.

AC AAK56009;

DT 31-AUG-1999 (first entry)

DE Merozoite surface protein MSP-1-42 cDNA.

KM MSP-1; merozoite surface protein; malaria; vaccine;

KW protein engineering; protein expression; codon usage;

OS Plasmidum falciparum.

XX Key Location/Qualifiers

FT CDS 1..1086

PN MO9920774-A2.

PD 29-APR-1999.

PF 20-OCT-1998; 98WO-US222226.

PR 15-MAY-1998; 98US-0085649.

PR 20-OCT-1997; 97US-0062592.

PA (GEN2) GENZYME TRANSGENICS CORP.

PI Chen LH, Meade H;

DR WPI: 1999-288313/24.

DR P-PSDB: AAY09373.

PS Modified malarial protein for use in anti-malarial vaccines

XX Example; Fig 2; 35pp; English.

XX This nucleic acid encodes a 42 kDa C-terminal portion (see AAY09373)

CC of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an

CC important target for the development of a vaccine against

CC Plasmodium falciparum. The 3' end of the sequence has been

CC modified to include a 6xHis tag. The nucleic acid sequence has

CC been modified (see AAK56008) according to a method of the invention

CC in order to improve expression in mammalian host cells and in

CC transgenic animals. In the modified coding sequence, 306 nucleotide

CC positions have been replaced to lower the AT content (from 76 to

CC 49.7%) and to eliminate 10 mRNA instability motifs. The encoded

CC amino acid sequence is unaltered. In another modified sequence

CC (see AAK56021), a signal peptide sequence has been added and two

CC N-glycosylation sites eliminated. The invention allows expression

CC of MSP-1-42 protein in the milk of transgenic animals, and also

CC provides a DNA vaccine comprising a vector containing the altered
 CC MSP-1-42 sequence.
 XX
 SO Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;

Query Match 12.7%; Score 625.2; DB 20; Length 1088;
 Best Local Similarity 74.3%; Pred. No. 6.6e-142;
 Matches 789; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Oy 3799 gctgtactctctccgtatgtataacatctgtccaaatcgsaagcagatg 3858
 Db 1 gaagtaactctctccgtatgtataacatctgtccaaatcgsaagcagatg 60
 Oy 3859 ctctatcgsaagcagatgataacatctgtccaaatcgsaagcagatg 3918
 Db 61 ttattatcaaaactcttgcagagtggttatagaagtttaaaaaaactatgaataaac 120
 Oy 3919 gtatgaccttcaatgtcaacgtgagacatctctgaacagcgcttataagagaa 3978
 Db 121 gtatgacatttaattgttaagttaagataattttaattcagatttaataacgtgaa 180
 Oy 3979 aatttcaagaacgtcttgagagcagacttgattccctataaagacgtacctctaac 4038
 Db 181 aatttcaaaaatgttttaagatcagatttaattccataaagaatttaacatcaagtaat 240
 Oy 4039 taagtgtcgaagacccataacatgtctcctataaagaagaaggataattctgtct 4058
 Db 241 tatgtgtcgaagaagatccataaaattcttataaagaagaagaagaataattcttaagc 300
 Oy 4099 agttacaactatcaagaagactccatcgacacgcagatacaatttcgtaagatgtgtc 4158
 Db 301 agtataatataatgaagatccaatagatacagataaataatttcgaatgagtctt 360
 Oy 4159 gggatatacaagaatccgtgagcaaaaatacaacagcttgacctgtactatataaagrat 4218
 Db 361 ggatattataaataatttccgaaaaataataatcagatttagaattcaatttaaaaaatat 420
 Oy 4219 atcaacgataagcagaagcgagagatgaanaatatctgcaccttcggaataacatcgaaac 4278
 Db 421 atcaacgataaacaagaagtgagaatgagaataactctcccttttaaaaaatattgagacc 480
 Oy 4279 ctgtcaagaagctgaaacgaacaaatcgacctctctgtcaattcactgagcgcaaggtc 4338
 Db 481 ttataataacagttatgataaataattgattatttgaattcatcttgaagcaaaagtt 540
 Oy 4339 ctcaactactactcagagaagaacaaatgtggaagttaaaatcaaggaagctgaactctc 4358
 Db 541 cttaattatacatatgagaataatcaacgtagaagtttaaaataaagaacttaattactta 600
 Oy 4399 aaacaatccaagaagcgtgagatttcaagaanaataatacaatttcgttgaattgca 4458
 Db 601 aaacaatccaagaagcgtgagatttcaagaanaataatacaatttcgttgaattgca 660
 Oy 4459 gactgtctacgattataacacacaacatctctgacacagttctgtccacgtgacatg 4518
 Db 661 gattatcaacagattataacacacaacatctctgacacagttctgtccacgtgacatg 720
 Oy 4519 gtttgcgaanaactcgcgaacaaacagctgtgagcaatctgtcgaagcagcagcagc 4578
 Db 721 gtttgcgaanaactcgcgaacaaacagctgttacttaacttactgttgcgaactgtgagc 780
 Oy 4579 atgtcgaacatctccgcgacacaaatggtgaagaanaagtgccccaagaatgagcgctg 4638
 Db 781 atgttaaacatcttcaacacacaaatgcttaaaaaaactgtccacaataatctgtgagat 840
 Oy 4639 ttcaagcactctgagcagcgacgaagtgcaagtctctccgaactatacaacaagaaga 4698
 Db 841 ttcaagcactctgagcagcgacgaagtgcaagtctctccgaactatacaacaagaaga 900
 Oy 4699 gataagtcgtgaggaacccaacctactctgaactgaagaatgaggtgtgaccc 4758
 Db 901 gataagtcgtgaggaacccaacctactctgaactgaagaatgaggtgtgaccc 960

Oy		4759	gattcctaatagcaccggcggaagacagcgcgtcttaaggagaazaatcacatgcagtgtc	4818
Dd		961	gatcccaacttaccgaaagaagatlccaagttagaacgaaaanaaatcatltylaatgt	1020
Oy		4819	actaagcccagaccttatccaccttcgaagcggatttttg	4860
Dd		1021	actaaccttgtattcttaccacttlctcgatgytatlttcltcg	1062
RESULT 14 AAx25587	ID	AAx25587	standard; cDNA; 1088 BP.	
XX	AC	AAx25587:		
XX	Df	02-AUG-1999	(first entry)	
XX	DE	Merozoite surface protein MSP-1-42	cDNA.	
KW	KM	MSP-1: merozoite surface protein; malaria; vaccine;		
KW	KM	protein engineering; protein expression; codon usage;		
xx	xx	transgenic animal; ss.		
OS	xx	Plasmodium falciparum.		
XX	FH	Key	Location/Qualifiers	
FT	CDS	1..1086		
FT		/+tag-	a	
PN		MO9920766-A2.		
PD		29-APR-1999.		
XX		20-OCT-1998;	98WO-US22225.	
PR		15-MAY-1998;	98US-0085649.	
PR		20-OCT-1997;	97US-0062592.	
PA		(GENZ) GENZYME TRANSGENICS CORP.		
P1		Chen LH, Meade H;		
DR		WPJ: 1999-302742/25.		
DR		P-PSDB; AAY05833.		
PT		New modified recombinant nucleic acid sequences useful for producing		
PT		malarial DNA vaccine		
XX		Disclosure; Fig 2; 43pp; English.		
CC		This nucleic acid encodes a 42 kDa C-terminal portion (see AAY05833)		
CC		of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an		
CC		important target for the development of a vaccine against		
CC		Plasmodium falciparum. The 3' end of the sequence has been		
CC		modified to include a 6His tag. The nucleic acid sequence has		
CC		been modified (see AAx25586) according to a method of the invention		
CC		in order to improve expression in mammalian host cells and in		
CC		transgenic animals. In the modified coding sequence, 306 nucleotide		
CC		positions have been replaced to lower the AT content (from 76 to		
CC		49.7%) and to eliminate 10 mRNA instability motifs. The encoded		
CC		nucleic acid sequence is unaltered. In another modified sequence		
CC		(see AAx25593), a signal peptide sequence has been added and two		
CC		N-glycosylation sites eliminated. The invention provides modified		
CC		recombinant nucleic acid sequences and methods for increasing the		
CC		mRNA levels and protein expression of proteins that are difficult		
CC		to express in cell culture systems, mammalian cell culture systems		
CC		or in transgenic animals. The preferred difficult protein		
CC		candidates for expression are those derived from lower organisms		
CC		such as parasites, bacteria and viruses that have DNA coding		
CC		sequences of high AT content or which have mRNA instability motifs		
CC		or rare codons relative to the recombinant expression system to be		
CC		used. The invention allows expression of MSP-1 protein in the milk		

CC	of transgenic animals, and also provides a DNA vaccine comprising a					
CC	vector containing the altered MSP-1-42 sequence.					
XX						
SQ	Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;					
	Query Match	12.7%;	Score 625.2;	DB 20;	Length 1088;	
	Best Local Similarity	74.3%;	Pred. No. 6, 6e-142;			
	Matches 789; Conservative	0;	Mismatches 273; Indels	0;	Gaps	
OY	3799 gctgcaccccttcgcatgtatgaatacattctctccaagaactcgagaaacagtg	3858				
Dd	1 gcagttaactccttcggcaattgtaaacatacttctaattggaatatgaatacagggt	60				
OY	3859 cctcatcgtgaaacctctgtaggagcgctctabgtctctcaagaacaagctgtagaatac	3918				
Dd	61 ttatatttaaacaccttagcagggtgtttatagagttttaaaaaaatltagaanaatac	120				
OY	3919 gtgatgacctcaatgtcacagtggaagagacattctgaacagcgctttaataagagaa	3978				
Dd	121 gttatgacatttaactgtlaatgtllaagatattaattcaattcaatlaataacgttaa	180				
OY	3979 aatttcaagaacgctctgtagagcgagactgattccctataaagagacctgcctcttac	4038				
Dd	181 aattcaaaaatgttttagatcagatuaattaccatataaagatttaacatcaagtaat	240				
OY	4039 taagtgltcaaaggaccatacaagttccctcaataaagaagaagatataattctgtc	4098				
Dd	241 tatgtgtcaaaagttccatbaataattctctaataaagaanaaagataattcttaagc	300				
OY	4099 agttacacataatcaaggagctccatccacgccgataataatttcgtaatgattgtgt	4158				
Dd	301 agttataatlatatlaagagatccaatagabaagatataaattctgcaatgagtctt	360				
OY	4159 gggtatlaacaagatccctgtagcgaaaaatacaaatctgacctgactctcataaagaat	4218				
Dd	361 ggaattatataataatataatccgnaaaaataaacgcgatttgatccaattaaanaaat	420				
OY	4219 atcaacagataagcaaggcgsgaaatgaaaaatalctgcctctccgtaatacatcgaaac	4278				
Dd	421 atcaacagcaaaacaagtgaaatgagaatatccctcccttttaaacataattgagac	480				
OY	4279 ctgtacagaagaatgaaacgcaaaatccgaccttctgtattccactgtaggcgaagtc	4338				
Dd	481 ttatatataaaacgctlaaagtataaactgtatttttgttaattcatattgaagcaaaagt	540				
OY	4339 ctoaacatacttaccggaagcgcaatgttggaagttaaancaagagctgtgacatcc	4398				
Dd	541 ctaaatcatatcatgtggaataatcaaacgctagaagaatgaataaagaacttattracta	600				
OY	4399 aaaacaatlccaagaacaagctgtgcagatttocaagaanaataaacaattctgcgaaatgca	4458				
Dd	601 aaacaattccaagaacaattgtgcagattttaaaaaaaaaataacaattctgttgaaatgct	660				
OY	4459 gaactgtctacagatttaaaccccacaatatctctgaccaagttctgtccactgtgcgt	4518				
Dd	661 gatattacaacagattataaocataaactatcttgacaagaattccttgaatcagattg	720				
OY	4519 gttgttgaaaaacctgcgcaaaacagatgctgtagaatactgtcgaagcgcaacttgaggc	4578				
Dd	721 gttcttgaaaactctgtctaaaaccttttatacttaattacttgtagyaaacttgcaaggt	780				
OY	4579 atgtctgacatctccccgacacatgtcgtagaanaacagtgcgcccaatatgcgagct	4638				
Dd	781 atgttaaacatttccaacacscatigtgtaaaaaaacaatgccaanaattctgtagt	840				
OY	4639 ttacagcatctygacgagcgcgagaagatgcgaagtgtctctcgaactacaacaagaagga	4698				
Dd	841 ttacagcatctttagttagaagaagagatgtaaatgtttatattcaacaacaagaagt	900				
OY	4699 gataaagtgcgtgagaacccaacacctactctgcaatgaaanaacaatgvgcggtgtgagcc	4758				
Dd	901 gataaagtgttgaanaatccaactctactctgttaacagaanaataaagtgtgagtatgca	960				

QY	4759	gdtgttaatagtcagcgaggaagacagcggtcttaacggaagaataatcatcgagtg	4818
DB	961		1020
QY	4819	actaagcccgactatccatctcttcgacgggattttgc	4860
DB	1021		1062
		actaaacctgattctatccacttttcggtgtattttctgc	
RESULT	15		
AAAC68978			
ACAC68978	standard: DNA: 786 BP.		
AC	ACAC68978;		
XX			
DT	27-FEB-2001 (first entry)		
XX			
DE	Merozoite surface protein-133 coding sequence.		
XX			
KM	Merozoite surface protein; protazoacide; vaccine; malaria; ss.		
XX			
OS	Plasmodium falciparum.		
XX			
PN	WO20063245-A2.		
XX			
PD	26-OCT-2000.		
XX			
PF	20-APR-2000; 2000WO-GB01558.		
XX			
PR	20-APR-1999; 98GB-0009072.		
XX			
PR	13-MAY-1999; 98US-0311817.		
XX			
PR	25-MAY-1999; 99CA-2271451.		
XX			
PA	(MED1-) MEDICAL RES COUNCIL.		
XX			
PI	Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiyibull C;		
XX			
DR	WPI: 2001-015762/02.		
XX			
DR	P-PSDB: AAB37610.		
XX			
PT	Novel variants of the C-terminal fragment of Plasmodium merozoite		
XX			
PT	surface protein-1, useful as vaccines for treating or preventing		
XX			
PT	malaria		
XX			
PS	Example 5; Fig 15; 126pp; English.		
XX			
CC	The present invention relates to non-natural variants of a C-terminal		
CC	fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The		
CC	non-natural variants have reduced affinity for at least 1 antibody		
CC	capable of blocking a second antibody that inhibits the proteolytic		
CC	cleavage of Plasmodium MSP-1 ₄₋₂ , and has the same affinity for at least		
CC	one third antibody that inhibits the proteolytic cleavage of Plasmodium		
CC	MSP-1 ₄₋₂ , compared to natural MSP-1 ₁₋₉ . The non-natural variants of the		
CC	present invention are useful for immunising a mammal against malaria, and		
CC	can be used to treat malaria. The present sequence is the MSP-133		
CC	coding sequence.		
XX			
SO	Sequence 786 BP: 245 A; 164 C; 159 G; 218 T; 0 other:		
Query Match	10.6%; Score 525.2; DB 22; Length 786;		
Best Local Similarity	79.3%; Pred. No. 1.2e-117;		
Matches	623; Conservative 0; Mismatches 163; Indels 0; Gaps 0;		
QY	3799	gctgtactctcttcggtgattgataacattctgtccaataatcgagaacgaatgagtg	3858
DB	1		60
		gctgtactctcatctgattatcgataacattctgtctaagattgagaacgaatacagagtc	
QY	3859	ctctatcgaaacctctgacgagcggtctatagtgcttccaagaacagctgagaataac	3918
DB	61		120
		ttgtactctgaagaccctctgacggtgctctacagatccctggaagaagcaactggaataaac	

[illegible]

PT New immunogenic polypeptides - derived from *Plasmodium falciparum* 190 kD
PT surface antigen precursor, useful in vaccines, and encoding DNA sequences
XX
PS Disclosure; : pp; German.

Query Match	6.3%	Score 308.8	DB 9	Length 660
Best Local Similarity	74.1%	Pred. No. 5.4e-65		
Matches 391	Conservative	0	Mismatches 137	Indels 0
				Gaps 0

QY	475	acactggtgtgtatcaactcaatcaggtcttcaaatctctgttgaagcttcgcgaagatcaat	534
Db	25	actctgtgtatcaatcattccatggtcttcaaatcttcaattgatcgtatctgagaatcaat	84
QY	535	gaactccgtcaacatcttgaaattcttacttcgactctgtctgaaggcgcaacsgaatcgtc	530
Db	85	gaattattatataaattcaaacctttatttgatttttaaggagaaaatttaaatgtia	144
QY	595	tgcgcgaatgactctgttcaaatccattcaatttgaagtcagcgcaacsgattgac	654
Db	145	tgtgtcatcgtctctgtccaatcccttccattcaaatctgtgcaaatgatctagac	204
QY	655	gtattgaagaagcttgcgtcttcgcgaatcgcgaagccttcgcgaacatcaaggacaattg	714
Db	205	gtacttaaaaactgtgttcgcgatagaaaaccattagacaatttaagaatgtia	264
QY	715	ggaagaatggaagtttatattaaaagaattaagaagccctcgtgaacatctaagctg	774
Db	265	ggaaaaatvgaaagcttccattcaaaaaataaaaaaccatagaaaataataatgaa	324
QY	775	atcgaagaatccaaaagaaccatagacatacaaaaataagaatgcacccaaggagaaag	834
Db	325	atggaagaagtaagaaaaacatctgtataaaataagaatgcacttaagaagaagaanaa	384
QY	835	aagaagcttgtaccgcggcccgtaagacctgtccatctatacaaaaacgcgttgaagaac	894
Db	385	aaaaaatataccaggtccatatagatcttcttatttaccataaaacattagaagaagca	444
QY	895	catacccatcagcgcgtcgtgaagaagcgcatagacaaccttcaagaagaatgaataatc	954
Db	445	catattttaataagcgttttagaaaaacgcatctgacacctttaaaaaaaatgtaaaacat	504
QY	955	aagaagactgctcgacaagatbaatgaaattaagaatccctccgcgagcc	1002
505	aaggaaattactctgataagataaataagaaatttaaaatcccccacgcgcg	552	

RESULT	18
AAAN81151	
ID	AAAN81151 standard; DNA; 1068 BP.
AC	AAAN81151;
DT	11-NOV-1990 (first entry)
XX	
DE	DNA sequence encoding polypeptide p190-3.
XX	
XX	Polypeptide p190-3; P.falciparum; metozoite; vaccine.
XX	
OS	Plasmodium falciparum.
XX	
XX	
FT	Key
FT	Location/Qualifiers
FT	1..1068.
FT	/tag= a
XX	
PN	EP283829-A.
XX	

PD	28-SEP-1988.	
XX		
PF	08-MAR-1988;	88BP-0103564.
XX		
PR	19-MAR-1987;	87GB-0006599.
XX		
PA	(HOFF) F. HOFFMANN-LA ROCHE & CO	

Query Match	6.3%	Score 308.8;	DB 9;	Length 1068;
Best Local Similarity	74.1%;	Pred. No. 6.9e-65;		
Matches 391; Conservative	0;	Mismatches 137;	Indels 0;	Gaps 0;

OY	475	agacgctgtgataaattccatcagcttcgaatatctgattgcggtttacgaagagataccat	533
Db	433	acttcgtgtgataaataattccatcaggtttccaaatattcattctgtgatatgagaatccaat	492
OY	535	gaactccgttacaagttaattgaatttctactgcgactgtctcaagggccaaacttgaacgtt	594
Db	493	gaattattatataaattaaacttattttgtattttatttaagagcaaatataatgtta	555
OY	555	tgcgcgaatgacattgttgcgaatcccatcatttgaagatcagaagccaaagattgac	655
Db	553	tgtgttaattgattatgttgcgaatcccttcaactcctaataattctgtgcaatgaattgac	612
OY	655	gtatcgaagaagtgtgctcttcgcgatatccgaagcgtctcgcgaacaatcaaggaacattgtg	714
Db	613	gtacttaaaaaaactgtgttcgcgatatagaagaacccattagccaattataagaataatgtta	672
OY	715	ggaagaatgtgagaattattataaagaataagaagaccatcgcgaacataacagactg	774
Db	673	ggaaaaaatgtgaagatttaccatttaaaaaataaaaaaacctagaaatatataatgaaatla	733
OY	775	atcgaagaatccaaaagaagaccattagacaaaataagaattgaacaccaaaggagatgaagaag	834
Db	733	attgtaagaagaattgaagaaacaatttgaataaataataagaattgacacttaagaagaagaaaaa	792
OY	835	aagaagctgttacccagccagttacgacctgttccattataaacaacagcttgaagaagcc	894
Db	793	aaaaattatataccaaagcccaatcagattctttttatttacaataaacatttgaagaagca	852
OY	895	cataaccctcatcagctgattctcgtggaagagcctagaagcccccagaagaatgaaatattc	954
Db	853	cataatttaataaagcgttttgaagaaacgcttttgacactttaaaaaaaatgtgaataacatt	912
OY	955	aagaagactgctgcgaacaagattatgtaaatataagaatccctccgcagcc	1002
Db	913	aaggaattactgtataagataataatgaaattaaataccccacagcgcc	960

RESULT	19
AAT80403	
ID	AAT80403 standard; DNA; 354 BP
XX	
AC	
AAT80403;	
XX	
DT	25-MAR-1998 (first entry)

XX	DE	XX	PFMSPI(p19)A coding sequence.
XX	XX	KW	Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX	KM	XX	Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
XX	OS	XX	Plasmodium falciparum.
XX	OS	XX	Synthetic.
XX	FT	FT	Key
XX	FT	FT	CDS
XX	FT	FT	Location/Qualifiers
XX	FT	FT	1..354
XX	FT	FT	/tag= a
XX	FT	FT	/product= "PFMSPI(p19)A with GPI anchoring sequence"
XX	FT	FT	/note= "sequence contains 2 in frame stop codons at the 3'-end"
XX	FT	FT	1..285
XX	FT	FT	/tag= b
XX	FT	FT	/note= "derived from P. falciparum MSP1 p19 sequence"
XX	FT	FT	286..354
XX	FT	FT	/tag= c
XX	FT	FT	/note= "glycosylphosphatidylinositol anchoring sequence"
XX	PN	XX	WO9730159-A2.
XX	XX	XX	21-AUG-1997.
XX	XX	XX	14-FEB-1997;
XX	XX	XX	97WO-FR00291.
XX	PR	XX	14-FEB-1996;
XX	XX	XX	96FR-0001821.
XX	PA	XX	(INSP) INST PASTEUR.
XX	PA	XX	(UNIV) UNIV NEW YORK STATE.
XX	PI	XX	Barnewell JW, Longacre-Andre S, Mendis K, Nato F;
XX	PI	XX	Roth C;
XX	DR	XX	WPI: 1997-425034/39.
XX	DR	XX	P-PSDB: AAM22592.
XX	XX	XX	Recombinant protein containing Plasmodium merozoite surface
XX	PT	XX	protein-1 p42 fragment - useful in antimalarial vaccines, also new
XX	PT	XX	antibodies for diagnosis and protein purification
XX	XX	XX	Disclosure: Fig 1B; 79pp: French.
XX	PS	XX	This is the nucleotide sequence encoding a recombinant protein comprising
XX	XX	XX	nucleotides encoding amino acids 1613-1705 of the Plasmodium falciparum
XX	CC	XX	merozoite surface protein 1 (MSP1) 19 kD C-terminal fragment (p19),
XX	CC	XX	linked to a sequence encoding a glycosylphosphatidylinositol membrane
XX	CC	XX	anchoring sequence. p19 is the C-terminal fragment of the 42 kD MSP1
XX	CC	XX	from Plasmodium species.
XX	CC	XX	The recombinant protein can be used for the production of anti-malarial
XX	CC	XX	vaccines, where the p19 fragment provides a high level of protective
XX	CC	XX	immunity since it includes epitopes not presented in the p42 fragment.
XX	XX	XX	Sequence 354 BP; 99 A; 92 C; 88 G; 75 T; 0 other;
XX	SQ	XX	
XX	XX	XX	Query Match 5.0%; Score 246.6; DB 18; Length 354;
XX	XX	XX	Best Local Similarity 81.7%; Pred. No. 5.4e-50;
XX	XX	XX	Matches 285; Conservative 0; Mismatches 64; Indels 0; Gaps 0
XX	QY	XX	4583 tgaacatccccaagcccaatgctgaagaacaacagtcgcccagaatagcgctgtttca 4642
XX	DB	XX	5 tcaaacatcccgagaccccaatgctgtaadaaaaaaacatgctccgagaactcgtgctgttca 64
XX	QY	XX	4643 ggcactctgacgacgcgcgaagctgaagctgttcctcgtgaactacaacaagaagagata 4702
XX	DB	XX	65 gacacttgtagcagagagagagtgtaaatgtctgctgaactacaacagagagcgaca 124
XX	QY	XX	4703 agtgcgtgtagaaccacaacctactctgcaatgaaacaacatgctgctgtgacgcgatg 4762
XX	DB	XX	125 agtgcgtgtagaaccacaacctactctgcaatgaaacaacatgctgctgtgacgcgatg 184

QY	4763	ctatctgaccacggagggaagcaggggctcctaacggagaagaataatcaatgcagatgta	4822
Db	185	ccaatgaccacggaggaggaactcgggcagacaagcgcaagaataacgcgtgagtgta	244
QY	4823	agcccgactctctatccactctctgcagcgggatttttgcctccagctctaattctcggca	4882
Db	245	aaccgcgactgtaaccgcgtgtcttgcagcggactctctgcagctcccttaactcttggca	304
QY	4883	tctctctcgtcgtgatccctcaatgcgtatccctgtacaagctcatcctaata	4931
Db	305	tctcgtctctgtgatccctcaatgattctgtatcgttaacagctcatcctaata	353
RESULT	20		
ID	AA794550	standard; DNN; 354	BP.
AC	AA794550;		
XX	AA794550;		
DT	25-MAR-1998	(first entry)	
XX	PI7MSPI(p19)A	coding sequence.	
DE			
XX	Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;		
KW	Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.		
XX	Plasmodium falciparum.		
OS	Synthetic.		
XX			
FT	Key	Location/Qualifiers	
FT	CDS	1..354	
FT	/tag- a	/product= "PI7MSPI(p19)A with GPI anchoring sequence"	
FT	/note= "sequence contains 2 in frame stop codons at the 3'-end"		
FT	misc-feature	1..285	
FT	/tag- b	/note= "derived from P. falciparum MSP1 p19 sequence"	
FT	misc-feature	286..354	
FT	/tag- c	/note= "glycosylphosphatidylinositol anchoring sequence"	
XX	W093730158-A2.		
XX	21-AUG-1997.		
XX	PD		
XX	PF	14-FEB-1997; 97WO-FR00290	
XX	PR	14-FEB-1996; 96FR-0001822	
XX	PA	(INSP) INST PASTEUR.	
XX	PA	(UYNY) UNIV NEM YORK STATE.	
XX	PI	Barnewell JW, Longacre-Andre S, Mendis K, Nato F, Roth C;	
XX	PI		
XX	DR	WPI: 1997-425033/39.	
XX	DR	P-PSDB; AAM36103.	
XX	PT	Recombinant protein containing the merozoite surface protein-1 p19	
XX	PT	fragment - useful in anti-malarial vaccines, diagnosis and protein	
XX	PS	purification	
XX	PS	Disclosure; Fig 1B: 85pp; French.	
XX	XX	This is the nucleotide sequence encoding a recombinant protein comprising	
XX	XX	nucleotides encoding amino acids 1613-1705 of the Plasmodium falciparum	
XX	XX	merozoite surface protein 1 (MSP1) 19 kD C-terminal fragment (p19),	
XX	XX	linked to a sequence encoding a glycosylphosphatidylinositol membrane	
XX	XX	anchoring sequence. p19 is the C-terminal fragment of the 42 kD MSP1	
XX	XX	from Plasmodium species.	
XX	XX	The recombinant protein can be used for the production of anti-malarial	

CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p12 fragment.
 XX
 SQ Sequence 354 BP; 99 A; 92 C; 88 G; 75 T; 0 other;

Query Match 5.0%; Score 246.6; DB 18; Length 354;
 Best Local Similarity 81.7%; Pred. No. 5.4e-50;
 Matches 285; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

OY 4583 tgaacatccagaccacatgctgtaagaacagctgccccagaatagcgctgttca 4642
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 5 tcaacatctcgcagcaccaatgctgtaaaaaacaatgcccagactctgctgttca 64
 OY 4643 ggcctctggagcagcggaagagtgcaagtgctctcctgaactaacaagaagagta 4702
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 65 gacacttggacgagagagagagtgtaaatgctcgtgaactacaagaagagcgca 124
 OY 4703 agtgcgtggaagaaacccaactctacatgaaacaagcggtgtgacccgagtg 4762
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 125 agtgcgtggaagaaacccaactctgtaacgagaacaagcggtgtgacgagagc 184
 OY 4763 ctaaatgcacgaggaagacagcgctcttaacggaagaataacatgagtgta 4822
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 185 ccaaatgcacgaggaagagactcgcgcagcaacggaagaataacatgagtgta 244
 OY 4823 agccgcactctacacactctctgacggagatttttgctccagctctaattctggca 4882
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 245 aaccgcactcgcacccgctgttcgacggcactctcgcagctcccttaactctggca 304
 OY 4883 tctctctcctgctatcctcatgctgctcctgtaacgctcatcata 4931
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 305 tctgctctctgtgctatcctcatgctgtaacgctcatcata 353

RESULT 21
 AAN50355
 ID AAN50355 standard; DNA; 737 BP.
 XX
 AC AAN50355;
 XX
 DT 22-OCT-1991 (first entry)
 XX
 DE Storage-specific, late schizont merozoite malaria antigen insert of
 DE plasmid pFCC-1.
 DE
 KW Malaria; vaccine; plasmid p31-1; ss.
 XX
 OS Plasmodium spp.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 2..737
 FT /*tag= a
 FT /label= malaria antigen
 XX
 PN W08503725-A.
 XX
 PD 29-AUG-1985.
 XX
 PF 20-FEB-1985; 85WO-GB00072.
 XX
 PR 20-FEB-1984; 84GB-0004378.
 XX
 PA (BIOJ) BIOGEN NV.
 XX
 PI Mach B, Perrin L, McGarvey M, Cheung A, Shaw A;
 XX
 DR WPI: 1985-223371/36.
 DR P-PSDB: AAP50304.
 XX
 PT Prod. of antigens of Plasmodium species, esp. of falciparum - by
 PT recombinant DNA methods giving polypeptide(s) for protecting
 PT against malaria or for diagnosis.

XX
 PS Disclosure; Fig 6; 49pp; English.
 XX

CC The DNA sequence encodes a Plasmodium falciparum, Plasmodium vivax,
 CC Plasmodium malariae and Plasmodium ovale antigen which may be used
 CC in the diagnosis of malaria and as a vaccine against malaria.
 CC Nucleotides 293-422 and 429-530 are used in the preparation of
 CC monoclonal antibodies against the malaria antigen.
 XX

SQ Sequence 737 BP; 253 A; 90 C; 124 G; 270 T; 0 other;

Query Match 4.9%; Score 243.4; DB 6; Length 737;
 Best Local Similarity 68.6%; Pred. No. 4.8e-49;
 Matches 359; Conservative 0; Mismatches 146; Indels 18; Gaps 1;

OY 10 atgaatacatctctctcctctgctatcttctgtttatcatcaataactcagtcgtg 69
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 225 atgaagatcatatctcttcttctgctatcttctgtttatcatcaataactcagtcgtg 284
 OY 70 acccagatctctcctcagagctgttgaagaactcggagcttggaaagatcgctcct 129
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 285 acacatgaagttatcaagaactcgttcaaaaactagaagctttagaagatgcagtatg 344
 OY 130 accgatacagctcgttccagaagagagatggtcgtatgaagaagagcagtgccagc 189
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 345 acaggtatagttatttcaaaagaaatgttatataatgaagagacagtggaaca 404
 OY 190 gccgttaacacagcaacacccggttctaagggtctgtgctagcgtgtgctccggtgg 249
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 405 gctgttaacaatagttacacccgttcaagtggttcaagttactcagtggtcagtggtg 464
 OY 250 tctgtgctcttgggggttcgtcgtccctccggcgagcggtgcagatcagtggtcagtg 309
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 465 tcaagtgctcagtgctcagtggtggtt-----tcaagtggtcagtggt 506
 OY 310 gcaagcggcgttcccggaagcagtcgaagaacacatccatctgcaaatcagcagatcc 369
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 507 gcttcaggttggttcaggttaattcaagaagcgtacaatccctcagataatcaagtgatca 566
 OY 370 gacgcgaagtcctacgcgcacccaagcagcagtggaagaatactctcctcatacaag 429
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 567 gatgtaaatcttaacgctgatttaaacatagagttcaaatctacttgctcattataa 626
 OY 430 gagctgaagatccacagttgttgcacctcctatactatgacgacactgtgataac 489
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 627 gaactcaaatatccgaactcttctgatttaacacatcatatgacttctgtgataat 686
 OY 490 attcatgcttcaaatatcgtatcagcgttcaagaagatca 532
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 687 attcatggttcaaatatttattatgagatagataagaagaatta 729

RESULT 22
 AAN82176
 ID AAN82176 standard; DNA; 668 BP.
 XX
 AC AAN82176;
 XX
 DT 12-DEC-1990 (first entry)
 XX
 DE 31-1 Repeated delete.
 XX
 KW Malaria; stage-specific late schizont-merozoite antigen 31-1;
 KW ss; vaccine; protozoan parasite; repitope.
 XX
 OS Plasmodium falciparum.
 XX
 PN EP254862-A.
 XX
 PD 03-FEB-1988.
 XX
 PF 20-JUN-1987; 87EP-0108867.

Db 20 acattgaagtgatgacacaaacattgcccaacacacatgcgtttaagaagcaatgtccacaaa 79
 QY 4628 atagcgctgtttcagcagcatctgacagcagcgcaagagtgtaagtgctccctgaactaca 4687
 Db 80 accccgagtgattcagacattcgtgacgagagagaagaatgtaagtgtctgtgaactaca 139
 QY 4688 aacaagaagagataagtcgctgaggaacccaacacctactctgaatgaagaacatgagcg 4747
 Db 140 agcagaagagtgatgtaagtgtgttgagaacccaacacctactctgaacgagaacagctg 199
 QY 4748 ggtgtatcccgctgtgtaataatgacacgaggaagacagcgctcttaacggaagaataaca 4807
 Db 200 gacgcagcgtcagcgtcgaatgacacgaagaagacctgtcttaacggaagaagatca 259
 QY 4808 catgcgagtgtaactaagcccgacatccatccactcttcgacgagatctttgtccagact 4867
 Db 260 ctgcggaatgtaactaagcccgacatccatcccttgttcgataatctctgtcttccct 319
 QY 4868 ctaattcctg 4879
 Db 320 ctaactaagtgg 331

RESULT 26

AAT80404
 ID AAT80404 standard; DNA; 387 BP.

XX AAT80404;

DT 25-MAR-1998 (first entry)

XX PFMSPI(p19)S coding sequence.

XX Chimeric: Plasmodium vivax; merozoite surface protein; MSP1; p19;

KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.

XX Chimeric - Plasmodium vivax.

OS Chimeric - Plasmodium falciparum.

XX Key Location/Qualifiers

FT CDS

FT 1..387

FT /tag= a PFMSPI(p19)S

FT /note= "sequence contains 2 in frame stop codons at the 3' end"

FT sig_peptide

FT 1..57

FT /tag= b

FT mat_peptide

FT 58..381

FT /tag= c

FT misc_feature

FT 1..96

FT /tag= d

FT /note= "sequence derived from P. vivax MSP1"

FT misc_feature

FT 97..102

FT /tag= e

FT /note= "sequence derived from generated restriction enzyme site"

FT misc_feature

FT 103..381

FT /tag= e

FT /note= "sequence derived from P. falciparum p19 coding sequence"

FT WO9730159-A2.

PN 21-AUG-1997.

PD 14-FEB-1997;

XX 97WO-FR00291.

PR 14-FEB-1996;

XX 96FR-0001821.

XX (INSP) INST PASTEUR.

PA (UYN) UNIV NEW YORK STATE.

XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;

PI Roth C;
 XX
 DR MPI; 1997-425034/39.
 DR P-PSDB; AAW22592.
 XX
 PT Recombinant protein containing Plasmodium merozoite surface
 PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
 PT antibodies for diagnosis and protein purification
 XX
 PS Disclosure; Fig 1C; 85pp; French.
 XX
 CC This is the nucleotide sequence encoding a chimeric protein comprising
 CC nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
 CC surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
 CC 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
 CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.
 XX
 SQ Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 other;

Query Match

Best Local Similarity 81.3%; Pred. No. 2.3e-37;
 Matches 226; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 4583 tgacatctccagacacaaatgcgtgaagaagaagtcgccccagaaatagcgtgttca 4642
 Db 101 tcaacatctcgcagcaccaatgctgaataaataaatgtcccgagactctgtgttca 160
 QY 4643 ggcattcgaagcagcgaagagtgcaagtgctctcctgaactacaagaagaagata 4702
 Db 161 gacacttgacgagagagagtgtaaatgtctgtgaactacaacagagagcgaca 220
 QY 4703 agtgcgtgagaaacccaacacctcctgaatgaataaataatgacggtgtgacgcgatg 4762
 Db 221 agtgcgtgagaaacccaacacctcctgaatgaataaataatgacggtgtgacgcgatg 280
 QY 4763 ctaatgcacccgagagaagcagcgtctcaacgagaagaataatcattcgaatgacta 4822
 Db 281 ccaaatgcacccgagagagtggtgcgacgaacgagaagaataatcattcgaatgacta 340
 QY 4823 agccgcactctatcactcttcgacgagatcttctgc 4860
 Db 341 aaccgcactctacccgctgttcgacgacatctctgc 378

RESULT 27

AAT94549
 ID AAT94549 standard; DNA; 387 BP.

XX AAT94549;

DT 25-MAR-1998 (first entry)

XX PFMSPI(p19)S coding sequence.

XX Chimeric: Plasmodium vivax; merozoite surface protein; MSP1; p19;

KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.

XX Chimeric - Plasmodium vivax.

OS Chimeric - Plasmodium falciparum.

XX Key Location/Qualifiers

FT CDS

FT 1..387

FT /tag= a

FT /product= PFMSPI(p19)S

FT /note= "sequence contains 2 in frame stop codons at the 3' end"

FT sig_peptide

FT 1..57

FT /tag= b

FT mat_peptide

FT 58..381

```

FT      misc_feature      /*tag- c
FT      1..96             /*tag- d
FT      /note- "sequence derived from P. vivax MSP1"
FT      misc_feature      /*tag- e
FT      97..102           /note- "sequence derived from generated restriction
FT      /note- "enzyme site"
FT      misc_feature      103..381
FT      /tag- e
FT      /note- "sequence derived from P. falciparum p19 coding
FT      sequence"
XX      WO9730158-A2.
XX      21-AUG-1997.
XX      14-FEB-1997; 97WO-FR00290.
XX      14-FEB-1996; 96FR-0001822.
XX      (INSP ) INST PASTEUR.
XX      (UNIT ) UNIT NEW YORK STATE.
XX      Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX      Roth C;
XX      WPI; 1997-425033/39.
XX      P-PSDB; AAM36102.
XX      Recombinant protein containing the merozoite surface protein-1 p19
XX      fragment - useful in anti-malarial vaccines, diagnosis and protein
XX      purification
XX      PS      Disclosure; Fig 1C; 85pp; French.
XX      This is the nucleotide sequence encoding a chimeric protein comprising
XX      nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
XX      surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
XX      19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
XX      p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
XX      The recombinant protein can be used for the production of anti-malarial
XX      vaccines, where the p19 fragment provides a high level of protective
XX      immunity since it includes epitopes not presented in the p42 fragment.
XX      SQ      Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 other:

```

```

Query Match      3.9%; Score 194.8; DB 18; Length 387;
Best Local Similarity 81.3%; Pred. No. 2.3e-37;
Matches 226; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```

```

QY      4583 tgaacatctccagccaatgctgaagaagaagtcgccccaagaatagcgctgttca 4642
DB      101 tcaacatctgcagcaccacatgctgaagaagaacatgtcccgagacactgtcgcttca 160
QY      4643 ggcactggaagcagcgcaagagtgctgctcctgaatacacaagaagaagata 4702
DB      161 gacacttgaagagagagagtgtaaatgtctgcgtgaactacaagaagagcgaca 220
QY      4703 agtgcgtggaagaccacaacctactctgaatagaacaatgagcggtgtgacgcgagt 4762
DB      221 agtgcgtggaagaccacaacctactctgaatagaacaagcggtgtgacgcgagt 280
QY      4763 ctaaatgacccaggaagacagcgctctaaggaagaataacatgcgagtacta 4822
DB      281 ccaaatgacccaggaagagactcgggcagcaagcaagaataacatcgtgtgagtgtaca 340
QY      4823 agccgcactatcacactcttcgaagcggttttttgc 4860
DB      341 aaccgcactgtaaccgcgtgttcgacgcatcttcgc 378

```

```

RESULT 28
AAD22459
ID      AAD22459 standard; DNA: 309 BP.
XX      AAD22459;
AC      AAD22459;
DT      12-FEB-2002 (first entry)
XX      Plasmodium falciparum PfMSP1.19 insert in pUC105-01 vector.
XX      DE      Plasmodium falciparum PfMSP1.19 insert in pUC105-01 vector.
XX      KW      pUC105-01 vector; immunostimulant; vaccine; immunisation; therapeutic;
XX      C3d; immune response; ds.
XX      OS      Plasmodium falciparum.
XX      PN      WO200177324-A1.
XX      PD      18-OCT-2001.
XX      PF      09-APR-2001; 2001WO-GB01599.
XX      PR      08-APR-2000; 2000GB-0008582.
XX      PA      (ADPR-) ADPROTECH LTD.
XX      PI      Steward M, Cox VF;
XX      DR      WPI; 2002-010909/01.
XX      PT      Novel variant DNA sequence useful in DNA vaccine, encodes naturally
XX      occurring protein and comprises a sequence non-identical to naturally
XX      occurring DNA sequence encoding the protein
XX      PS      Example 13; Page 65; 87pp; English.
XX      CC      The invention relates to a variant DNA sequence useful in DNA vaccines.
XX      CC      The DNA sequence encodes a naturally occurring protein such as C3d which,
XX      CC      by virtue of third base redundancy and other variations permissible
XX      CC      within an amino acid codon, is non-identical to the naturally occurring
XX      CC      DNA sequence encoding that protein. The DNA sequence is useful in a DNA
XX      CC      immunisation vector to encode one or more naturally occurring human or
XX      CC      non-human proteins with immunomodulatory properties. The DNA sequence
XX      CC      is useful for inducing an immune response to an antigen in a human or
XX      CC      animal. A pharmaceutical composition is useful for introducing a DNA
XX      CC      sequence encoding a naturally occurring protein into a human or animal,
XX      CC      by administering the pharmaceutical composition into the human or animal,
XX      CC      where the administration results in a therapeutic effect on the human or
XX      CC      animal. The present sequence is Plasmodium falciparum PfMSP1.19 insert
XX      CC      in pUC105-01 vector
XX      SQ      Sequence 309 BP; 94 A; 71 C; 69 G; 75 T; 0 other:

```

```

Query Match      3.9%; Score 194.6; DB 24; Length 309;
Best Local Similarity 79.6%; Pred. No. 2.3e-37;
Matches 230; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```

```

QY      4585 aaactctccagcccaatgctgaagaagaagtcgccccaagaatacgctgttcaag 4644
DB      7 aacatgcccacaacccaatgctgaagaagaagtcgccacaacatccggtgttca 66
QY      4645 catctggaagcgcggaagagtgctgctcctgaatacacaagaagaagataag 4704
DB      67 catctggaagcgcggaagagtgctgctcctgaatacacaagaagaagtgataag 126
QY      4705 tgcgtggaagaccacaacctactctgaatagaacaatgagcggtgtgacgcgagt 4764
DB      127 tgcgtggaagaccacaacctactctgaatacagaagaacggtgtgacgcgagt 186
QY      4765 aaatgacccaggaagacagcgctctaaggaagaataacatgcggtgtactaag 4824
DB      187 aagtgcacccgaagaagactgtgttctaagcgaagaagaatattactgtgagtactaag 246

```


PT Novel variant DNA sequence useful in DNA vaccine, encodes naturally
 PT occurring protein and comprises a sequence non-identical to naturally
 PT occurring DNA sequence encoding the protein -
 XX
 PS Claim 16: Page 69-71; 87pp; English.

CC The invention relates to a variant DNA sequence useful in DNA vaccines.
 CC The DNA sequence encodes a naturally occurring protein such as C3d which,
 CC by virtue of third base redundancy and other variations permissible
 CC within an amino acid codon, is non-identical to the naturally occurring
 CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
 CC immunisation vector to encode one or more naturally occurring human or
 CC non-human proteins with immunomodulatory properties. The DNA sequence
 CC is useful for inducing an immune response to an antigen in a human or
 CC animal. A pharmaceutical composition is useful for introducing a DNA
 CC sequence encoding a naturally occurring protein into a human or animal
 CC by administering the pharmaceutical composition into the human or animal,
 CC where the administration results in a therapeutic effect on the human or
 CC animal. The present sequence is Plasmidium falciparum PfMSPI.19-human
 CC C3d3 DNA coding sequence from vector pVK104-02.

XX
 SO Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other:

Query Match 3.9%; Score 194.6; DB 24; Length 3147;
 Best Local Similarity 79.6%; Pred. No. 7.6e-37;
 Matches 230; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
 QY 4585 aacatctccagcaccatgctgagaagaacagtgccccagaatagcggtgttcag 4644
 Db 2845 aacatctccagcaccatgctgagaagaacagtgccccagaatagcggtgttcag 2904
 QY 4645 catctgagcagcgcggaagatgcaagtgtctcctgaactacaagaagaagataag 4704
 Db 2905 catctgagcagcggaagatgcaagtgtctcctgaactacaagaagaagataag 2964
 QY 4705 tgcgtgagaaacccaacctactctgcaatgaaacaatgycgggtgtgacgccgagtct 4764
 Db 2965 tgcgtgagaaacccaacctactctgcaatgaaacaatgycgggtgtgacgccgagtct 3024
 QY 4765 aaatgacccgaggaagacagcggtcttaacggaagaataacacatcgagtgtactaag 4824
 Db 3025 aaatgacccgaggaagacagcggtcttaacggaagaataacacatcgagtgtactaag 3084
 QY 4825 cccgactctctcactcttcgacggaatttttgctcagcttaatt 4873
 Db 3085 cccgactctcactcttcgacggaatttttgctcagcttaatt 3133

Search completed: September 20, 2002, 06:51:59
 Job time: 23633 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 00:16:46 ; Search time 126.47 Seconds
(without alignments)
9594.604 Million cell updates/sec

Title: US-09-269-874A-2
Perfect score: 4940
Sequence: 1 cgcacgcgtatgaataatcat.....ttcatctaatagatgatg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_MA: *
1: /cgn2_6/p/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/p/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/p/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/p/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/p/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/p/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1265.2	25.6	5181	1	US-08-257-073-10 Sequence 10, Appl
2	74.2	1.5	7218	1	US-08-232-463-14 Sequence 14, Appl
3	60.6	1.2	5163	3	US-08-700-651-1 Sequence 1, Appl
4	60.6	1.2	5163	3	US-08-928-361B-4 Sequence 4, Appl
5	60.6	1.2	5318	3	US-08-700-651-2 Sequence 2, Appl
6	60.6	1.2	5318	3	US-08-928-361B-3 Sequence 3, Appl
7	58.8	1.2	5311	3	US-08-928-361B-2 Sequence 2, Appl
8	58.8	1.2	7334	3	US-08-928-361B-1 Sequence 1, Appl
9	50	1.0	5661	4	US-08-938-105-2 Sequence 2, Appl
10	49.6	1.0	1430	2	US-08-276-452A-25 Sequence 25, Appl
11	48.2	1.0	1430	2	US-08-798-744-25 Sequence 24, Appl
12	48.2	1.0	1690	2	US-08-798-744-24 Sequence 24, Appl
13	48.2	1.0	1690	2	US-08-798-744-24 Sequence 24, Appl
14	46.4	0.9	198	5	PCT-US95-10668-3 Sequence 3, Appl
15	46.4	0.9	198	5	PCT-US95-10668-4 Sequence 4, Appl
16	46.4	0.9	954	4	US-08-098-327E-37 Sequence 37, Appl
17	46.4	0.9	954	4	US-08-462-625-37 Sequence 37, Appl
18	46.4	0.9	988	4	US-08-098-327E-34 Sequence 34, Appl
19	46.4	0.9	988	4	US-08-462-625-34 Sequence 34, Appl
20	45.8	0.9	340	1	US-08-182-175A-104 Sequence 104, App
21	45.8	0.9	340	5	PCT-US92-06412-104 Sequence 104, App
22	45.6	0.9	697	6	5171843-10 Patent No. 5171843
23	45.6	0.9	1137	6	5171843-8 Patent No. 5171843
24	45.4	0.9	198	5	PCT-US95-10668-1 Sequence 1, Appl
25	45.4	0.9	198	5	PCT-US95-10668-2 Sequence 2, Appl
26	45	0.9	2949	4	US-09-412-554A-3 Sequence 3, Appl
27	44.8	0.9	950	4	US-08-098-327E-32 Sequence 32, Appl

28	44.8	0.9	950	4	US-08-462-625-32 Sequence 32, Appl
29	44.8	0.9	1482	4	US-08-098-327E-41 Sequence 41, Appl
30	44.8	0.9	1482	4	US-08-098-327E-45 Sequence 45, Appl
31	44.8	0.9	1482	4	US-08-462-625-41 Sequence 41, Appl
32	44.8	0.9	1482	4	US-08-462-625-45 Sequence 45, Appl
33	44.8	0.9	2223	4	US-08-257-073-4 Sequence 4, Appl
34	44.4	0.9	4766	5	PCT-US93-07261-10 Sequence 10, Appl
35	44.2	0.9	2277	1	US-08-676-967-2 Sequence 2, Appl
36	44.2	0.9	2277	1	US-08-676-967-2 Sequence 2, Appl
37	44.2	0.9	2277	2	US-09-098-487-2 Sequence 2, Appl
38	43.6	0.9	1956	4	US-08-559-896B-1 Sequence 1, Appl
39	43	0.9	1241	1	US-08-471-033-39 Sequence 39, Appl
40	43	0.9	1241	1	US-08-471-033-42 Sequence 42, Appl
41	43	0.9	1241	2	US-08-471-044-39 Sequence 39, Appl
42	43	0.9	1241	2	US-08-471-044-42 Sequence 42, Appl
43	43	0.9	1241	2	US-08-463-483A-39 Sequence 39, Appl
44	43	0.9	1241	2	US-08-463-483A-42 Sequence 42, Appl
45	43	0.9	1241	2	US-08-471-046A-39 Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-08-257-073-10
Sequence 10, Application US/08257073
Patent No. 5765597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CUMTMS
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 5181 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-10

[illegible]

Db 4249 TCATTTAAATTTATTTGAATTCAGAACAAAAACACCTTTTAAAGTTACAAATATATA 4308
 Qy 4114 aagagctccatcgacacccatataatctgcaatgctgctggtatatacaagatc 4173
 Db 4309 AAGAGTCAAGTAAATATGATTTAATTTGACAGGAAGTATAGTTATTTGAAAAG 4368
 Qy 4174 ctgagcgaaataatacaagcttgactctataaagaatatacaagataag--- 4230
 Db 4369 GTTTTTCGCAAAATTAAGATGATTTAGATTAATTAATAAAGTTATCAAAAGAAAG 4428
 Qy 4231 -----caaa 4233
 Db 4429 GAGAGTTCCCATCATCACCACCAACACCTCCCTCACCAACAAAACAGACGACAA 4488
 Qy 4234 ggcgaatgaaataataatctgccccttcgtaataacatcgaaacccgtacaagagtg 4293
 Db 4489 AAGAGGAAAGTAAAGTTCCCTCATTTTAAACAAACATTGAGACCTTATACAACTTA 4548
 Qy 4294 aacgacaaatcgacctctcgtaattcaccctgagagcgaagctcctcaactaac 4353
 Db 4549 GTTAATTAATTAATGACATTAATTAATTAATTAATTAATTAATTAATTAATGTT 4608
 Qy 4354 gagaagaagaatctggaatgaaatcaagaagctgaactcctcaaaacatccaagac 4413
 Db 4609 GAAAAAGATGAAACACATGTTAAATTAATTAATTAATTAATTAATTAATTAATGAC 4668
 Qy 4414 aagctgagcagattcgaagaataatacaatctgctgcaatctgcaagcctgctaccg 4473
 Db 4669 AAAAAATGATCTTTTAAAAACCATTAAGACTTGAAGCAATTAATAAATGAT 4728
 Qy 4474 tataaccaacaacatctcctcgacgaagttctgcaactgctgctgctgctgctgaac 4533
 Db 4729 GATACGAAAAAAGATATGCTTGGCAATTAATTAATTAATTAATTAATTAATTTT 4785
 Qy 4534 gccaagaagctgctgagcaatctgctgcaagcgaagcctgcaagcctgcaagcctc 4593
 Db 4786 CCAATATACATATATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTTCA 4845
 Qy 4594 cagcacaatctgctgaagaacagctgcccagaatagcgtctgttcagcactgagac 4653
 Db 4846 CAACAGCAATGCGTAAATAAACAATGTCGAAATTTCTGATGTTTACAGATTTAGAT 4905
 Qy 4654 gagcggaagatgcaagtgctcctcgactaacaagaagaagataagtgctgctgag 4713
 Db 4906 GAAAGGAAAGAAATGTAATTTATTAATTAATTAATTAATTAATTAATTAATTTGAA 4965
 Qy 4714 aaccacaacccctcctgcaatgaagaacatgctgctgctgctgctgctgctgctgct 4773
 Db 4966 AATCCAAATCTACTTGTAAACGAAATTAATGATGATGATGATGATGATGATGATGATG 5025
 Qy 4774 gaggaagaagcgtctcaacggaagaataacatctgcaagtgctgctgctgctgctgct 4833
 Db 5026 GAAAGAGATTCAGTGAAGCAAGAAAGAAATACATGTAATGTAATGTAATGTAATGTA 5085
 Qy 4834 tatcaactctgagagagatcttctgctcagctcaactctgctgctgctgctgctgct 4893
 Db 5086 TATCCACTTTTCATGATGATTTTCTCAGTTCTCTTAACCTTTTGAATATCATCTCTTA 5145
 Qy 4894 ctgactcctatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 4929
 Db 5146 TTAATCTCATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5181

RESULT 2

US-08-232-463-14/c
 ; Sequence 14, Application US/08232463
 ; Patent No. 3670367
 ; GENERAL INFORMATION:
 ; APPLICANT: DORNER, F.
 ; APPLICANT: SCHEIFLINGER, F.
 ; APPLICANT: FALKNER, F. G.
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 ; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 1800 Diagonal Road, Suite 500
 ; CITY: Alexandria
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22313-0299
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/232,463
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/935,313
 ; FILING DATE:
 ; APPLICATION NUMBER: EP 91 114 300.6
 ; FILING DATE: 26-AUG-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703)836-9300
 ; TELEFAX: (703)683-4109
 ; TELEX: 899149
 ; INFORMATION FOR SEQ. ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 7218 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; CLONE: PTZgpt-F15
 ; US-08-232-463-14

Query Match 1.5%; Score 74.2; DB 1; Length 7218;
 Best Local Similarity 7.8%; Pred. No. 2.5e-09;
 Matches 34; Conservative 224; Mismatches 167; Indels 0; Gaps 0;
 Qy 1693 aaagaactaagatatacaagaatctcatagtaagatcgaagaagatcgagcgt 1752
 Db 1465 AAGTACTTAAGATGAAGAAAGATTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRR 1406
 Qy 1753 gtgagaacatlaagaagatgagaacagtgcttgagaagaagattacaagaagcga 1812
 Db 1405 RRR 1346
 Qy 1813 aatlaaccagatlaagaagatcccgagtgctcgatattgaatcgaagtcgaag 1872
 Db 1345 RRR 1286
 Qy 1873 gtgctcctcatagaagaatgtaagaactcaagaactcaactctctgaagaagcgt 1932
 Db 1285 RRR 1226
 Qy 1933 gagtlaaacaataatatactgctcggaatagttacaagcagaagataagcagaacca 1992
 Db 1225 RRR 1166
 Qy 1993 tactactctatgtaagaagaagatgagaacatgaagtgctatgcccagaagtc 2052
 Db 1165 RRR 1106
 Qy 2053 gagaactgtaacaagaagaagaacatcaaaactgaagaagtcgaagtcgaactcc 2112
 Db 1105 RRR 1046
 Qy 2113 gagccttcacagaa 2127

Db 1045 CAGCCAAGCTCGAA 1031

```
RESULT 3
US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEBCH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700.651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5163
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-1
```

Query Match 1.2%; Score 60.6; DB 3; Length 5163;
Best Local Similarity 44.2%; Pred. No. 8.2e-06;
Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

```
QY 3930 caatgtcaagtgtagagacattctgacacgcgcttaataagagagaatttcaaga 3989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 caacaacaacaacaacaacaacaacaacaacaacaacaacaacaacaacaaca 674
QY 3990 cgtcttgagagcgactgattccctataaagacgtgacccctctaactcgtgtcaa 4049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 caacaacaacaacaacaacaacaacaacaacaacaacaacaacaacaacaaca 734
QY 4050 ggaaccatacaagttcctcaataaagaaggatgaatttctgtctagttacaata 4109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 735 cgactactactactactactactactactactactactactactactactactaa 794
QY 4110 tatcaaggactccatcgacaccgatataattcgctaatatgtgtcgtgggtataaa 4169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 795 caaccacaactacaacaacaactacaaccacaacaacaacaacaacaactacaa 854
QY 4170 gatcttgagcgaaataatcaagctgactgactctataaaagtatatcaacgataa 4229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 855 ccaacaactacaaccacaacaactacaaccacaacaacaacaacaactacaa 914
QY 4230 gcaagcgagaaatgaanaatctgccccttcgtgaataacatcgaaacctgtacaagc 4289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 915 agaacaacaacaacaactacaacaacaacaacaacaacaactactactactaa 974
QY 4290 agtgaagacaanaatgcacctcttcgttaattcactcgtgagcgcaaggtcccaactaac 4349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 975 ccaacaacaacaacaactacaactacaactacaagaagaacaaactactactacta 1034
QY 4350 ttacgagaagcgaaatgtggaagttaaatcaaggagctgactaccccaaaacatcca 4409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035 ctacaacaacaacaactactactactactacacaacaacaacaactactactacta 1094
QY 4410 agacaagctgcaagatttcaagaaanaatacaatttcgtcggaattgagacgtgtcac 4469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1095 caacaacaacaacaacaacaacaacaacaactacacgaaacaaacaacaacaaca 1154
QY 4470 cgattataaccacaacatctcc 4492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1155 caactactactactactaacac 1177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 4
US-08-928-361B-4
; Sequence 4, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:

```
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928.361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: VERNY, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5163 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-4
```

Query Match 1.2%; Score 60.6; DB 3; Length 5163;
Best Local Similarity 44.2%; Pred. No. 8.2e-06;
Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

```
QY 3930 caatgtcaagtgtagagacattctgacacgcgcttaataagagagaatttcaaga 3989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 614 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 673
QY 3990 cgtcttgagagcgactgattccctataaagacgtgacccctctaactcgtgtcaa 4049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 733
QY 4050 ggaaccatacaagttcctcaataaagaaggatgaatttctgtctagttacaata 4109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 734 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 793
QY 4110 tatcaaggactccatcgacaccgatataattcgctaatatgtgtcgtgggtataaa 4169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 794 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 853
QY 4170 gatcttgagcgaaataatcaagctgactgactctataaaagtatatcaacgataa 4229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 854 CCACAACACTACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


QY 3990 cgttttggaagcacttgccttaaaaggacgccccttcctaagtgtgca 4049
| | | | |
Db 674 CAACAACAAACAGAACACCAACAACAACAACAACAACAACAACAACA 733

QY 4050 ggaccatacaagtlccctccaatlaagaagaaggatnaatttcgttagttaacta 4109
| | | | |
Db 734 CGACTTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA .793

QY 4110 tatcaaggacctcatcgacaecggatalcaatlctgctaagtgtgtcggggtatta 4169
| | | | |
Db 794 CAAACCACAACTTAACACCCACAACAATAACACCAAACAACAACAACAAC 853

QY 4170 gatccggaggaaaatatcaagltcgtacctgatctcataaaaaagtatcaaatga 4229
| | | | |
Db 854 CCACAACTATCACACCACCAACAACTTAACACCAACAACCAACCAACCAAC 913

QY 4230 gcaaggcgagatlgaaaatatctgcccttcctcgtgaatacacatgaaacctglta 4289
| | | | |
Db 914 AGAAAACAACAACAAACAACAACTTAAAAACAACAACAACAACAACAACA 973

QY 4290 agtgaagcgcaaataatcgacctcttcgtgaatlcacctggaagcgaaagtcctcaata 4349
| | | | |
Db 974 CCACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1033

QY 4350 ttacggaagaagcaatyltggaagtltaaatacaaggagctlgaactaccaccaaacatca 4409
| | | | |
Db 1034 CTATCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1093

QY 4410 agacaaagctggcagatlltcaagaaaataacaatlctgvggaatlbgagacctgttac 4469
| | | | |
Db 1094 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1153

QY 4470 cgatlaaacacacaactctcc 4492
| | | | |
Db 1154 CAACTACTACTACTACTACTAACC 1176

RESULT 7
US-08-928-361B-2
; Sequence 2, Application US/08928361B
; Patent No. 6071518

; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTES;
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESSES:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840

; COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928, 361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026, 062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30, 518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:

```

: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 5511 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-928-361B-2

Query Match      1.28: Score 58.8; DB 3: Length 5511;
Best Local Similarity 45.28: Pred. No. 2.6e-05;
Matches 216: Conservative 0; Mismatches 262; Indels 0; Gaps 0;

OY 3936 caacgtgaagacaccttcgaacagcgctttaaagaagaaattccaagacgtcctt 3995
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1094 CCAGCAGACACTACACCAACCAACCAACCACTACCAAGAAACCAACACACCAACCA 1153
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

OY 3996 ggaagagcaacttgattccctataaagaacctgaacctccttaactgaagtgtcaagacc 4055
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1154 CAACACTACTACTACTACCAACCAACCAACCAACCAACCACTACTACTACTACTACTACTA 1213
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

OY 4056 atacaagtccctaataaagaagaggaatatttcgtctgaagtaacacatatacga 4115
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1214 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1273
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

OY 4116 ggaactcatcgacacgcgaatcaatttcgcctaagtgtggtggtatatacagaatccct 4175
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1274 CAACACACACACACACACACACACACACACACACACACACACACACACACACACACACGA 1333
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

OY 4176 gagcgaaataacagaagctcgaccttgacctctattaaagaatatacaagaataagcaagg 4235
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1334 CAACACACACACACACACACACACACACACACACACACACACACACACACACACACAA 1393
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

OY 4236 cgagaatgaanaatatctgccttcctcgaataacatcgaaacctgtacaagaacgtgaa 4295
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1394 CAACACACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1453
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

OY 4296 cgacaaatcgacctcttcgttaattcacctcgagagccaaagctcctaactatacttaga 4355
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1454 CCACAGACACCAACCGCAACCAACCACTACCAAGAAACCAACACACACACACACACACTACTA 1513
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

OY 4356 gaagagcaatgtggaagtaaataacaaagagctgaactactcctaanaacatccaagac 4413
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 1514 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1571
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT      8
US-08-928-361B-1
: Sequence 1, Application US/08928361B
: Patent No. 6071518
: GENERAL INFORMATION:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: TITLE OF INVENTION: SPECIES INFECTIONS
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306-1840
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/928.361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, HANA
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-928-361B-1

```

```

Query Match          1.2%; Score 58.8; DB 3; Length 7334;
Best Local Similarity 45.2%; Pred. No. 3e-05;
Matches 216; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

```

```

QY 3936 caagcgtgaagagcattctgacacagccgtttaaagaagaataattcaagaagctctt 3995
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2762 CCACGACAACTACACACACACACACAACTACCAAGAAACCAACACACACACAAACAA 2821
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3996 ggaagagcacttgattccctataaagaacctgacctcccttaactaactggtgcaagacc 4055
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2822 CAACCTACTACTACTACTACAAACAAACAAACAACTACTACTACTACTACTACTACTACT 2881
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4056 atacaagttccctaaataaagaagaaggataattctgtctgaagtaacactatctaa 4115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2882 CTACTACTACTACTACTACTACACAAACCAACCAACCAACCAACCAACCAACCAACCA 2941
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4116 ggaactcagcagcaccgatacaatttcgttaagtgtggtggtatatacaagctct 4175
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2942 CAACACAAACAAACAAACAAACAAACAACTACTACTACTACTACTACTACTACTACTACT 3001
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4176 gaagcgaataatacaagctgacctgacctatataaagaatatacaagaataagaag 4235
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3002 CAACCAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 3061
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4236 cgagaatgaaaaatattcgtccctctgataaatacgcgaaccctgtacaagaagctgaa 4295
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3062 CAACCAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4296 cgacaaatcgacctctcgttaattcaactgagcgaagctcctcaactataactacga 4355
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3122 CCACAAACCAACACCGACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 3181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4356 gaagagcaatgtagaagttaaatacaagaagctgaactactactaaatacaatccaagac 4413
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3182 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3239
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 9
; US-08-938-105-2
; Sequence 2, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Leinwand, Leslie A.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver

```

```

; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,105
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crook, Wanneil M.
; REGISTRATION NUMBER: 31,071
; REFERENCE/DOCKET NUMBER: 3595-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5661 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..5661
; US-08-938-105-2

```

```

Query Match          1.0%; Score 50; DB 4; Length 5661;
Best Local Similarity 46.3%; Pred. No. 0.0056;
Matches 164; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

```

```

QY 738 aaagaataagaagaaccatcgagaaacttaacgagctcgatcgaaatccaaaagaaccat 797
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2376 AGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2435
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 798 agacaaaataaagaatcgcacccaagaagaagaagaagaagtgtacacagccagta 857
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2436 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2495
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 858 cgacctgcatcattataaanaacagctgaagaagcccaataacccaatcaagctactgga 917
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2496 GAAGATGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 2555
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 918 gaagcgatagaaacctctaaagaagaatgaataatatacaagaactctcgacaagttaa 977
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2556 GCGCTGCGACCACTGATCAAGAACCAAGATCCAGTGGAGGCCAAGGTGAAGAGATGAG 2615
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 978 tgaattaaagaatccctccgcgaaccaactctgggaacacccctaaacgctgtgagacaa 1037
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2616 CGAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2675
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1038 gaacaaagaatagaggagcagagaagaagatcaaaagagatcgccaataaccat 1091
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2676 GGAAGACGAGTGTCTGAGAGCTCAAGAAAGATATGATGATGATGATGATGATGATGAT 2729
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 10
; US-08-276-452A-25
; Sequence 25, Application US/08276452A
; Patent No. 5646029
; GENERAL INFORMATION:
; APPLICANT: Chen, Chao-Guang
; TITLE OF INVENTION:
; APPLICANT: Mau, Shiao-Lim
; APPLICANT: Du, He
; APPLICANT: Gane, Allison M
; APPLICANT: Bacik, Antony
; APPLICANT: Clarke, Adrienne E
; TITLE OF INVENTION: Plant Arabidnoglactan Protein (Acp) Genes

```



```

: NUMBER OF SEQUENCES: 91
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee and Winner, P.C.
: STREET: 5370 Manhattan Circle, Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: United States of America
: ZIP: 80303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/276,452A
: FILING DATE: 18-JUL-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Caruthers, Jennie M.
: REGISTRATION NUMBER: 34,464
: REFERENCE/DOCKET NUMBER: 27-91A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303)499-8080
: TELEFAX: (303)499-8089
: TELEX: 49617824
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1430 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..1312
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 41..112
: OTHER INFORMATION: /note= "Derived amino acid sequence
: OTHER INFORMATION: corresponding to the peptide sequence by protein
: OTHER INFORMATION: microsequencing"
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 41..112
: OTHER INFORMATION: /note= "Derived amino acids 14-24,
: OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
: OTHER INFORMATION: by protein microsequencing"
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 25..31
: OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
: OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
: OTHER INFORMATION: 26 can be T instead of A"
: US-08-276-452A-25

Query Match 1.0%; Score 49.6; DB 1; Length 1430;
Best Local Similarity 44.3%; Pred. No. 0.0031;
Matches 202; Conservative 0; Mismatches 254; Indels 0; Gaps 0;

```

```

: 3085 ggcaagataaagatgagatgaagaagtgtgctcgaagagagagcttgaagcaaa 3144
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 521 GGCTACTCGGAGATATACGACAAACAAACAAATGCTACTCGAGAAATACAAAC 580
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: QY 3145 ctcaactcactgaacaatcgaacaacgtaactgcagaactctcagtgltctcaacaag 3204
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 581 AACCAACACAAATGCTACTCCGAGAAATTACAAACAAACAAATGCTCAAGAAT 640
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: QY 3205 aagaaggaagccgagatgcgcgagagacgaacactctgagagaaacgaattcttc 3264
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: DB 641 TACACACAAATGCTACTCCGAGAAATGCTACTCCGAGAAATACAAATGCTCAAG 700
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
: QY 3265 aaacactacaagcctcgtcaagtattataatgctgagcttctcctcgtgaactctc 3324
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-798-744-25
: Sequence 25, Application US/08798744
: Patent No. 5830747
: GENERAL INFORMATION:
: APPLICANT: Chen, Chao-Guang
: APPLICANT: Meu, Shao-Lim
: APPLICANT: Du, He
: APPLICANT: Gane, Alison M
: APPLICANT: Bacic, Antony
: APPLICANT: Clarke, Adrienne E
: TITLE OF INVENTION: Plant Arabidogalactan Protein (AGP) Genes
: NUMBER OF SEQUENCES: 91
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee and Winner, P.C.
: STREET: 5370 Manhattan Circle, Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: United States of America
: ZIP: 80303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/798,744
: FILING DATE: 13-FEB-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/276,452
: FILING DATE: 18-JUL-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Caruthers, Jennie M.
: REGISTRATION NUMBER: 34,464
: REFERENCE/DOCKET NUMBER: 27-91A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303)499-8080
: TELEFAX: (303)499-8089
: TELEX: 49617824
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1430 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..1312
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 41..112

```

OTHER INFORMATION: /note- "Derived amino acid sequence
OTHER INFORMATION: corresponding to the peptide sequence by protein
OTHER INFORMATION: microsequencing"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 41..112
OTHER INFORMATION: /note- "Derived amino acids 14-24,
OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
OTHER INFORMATION: by protein microsequencing"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 25..31
OTHER INFORMATION: /note- "Amino acids 25, 27, 29, and
OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
OTHER INFORMATION: 26 can be T instead of A"
US-08-798-744-25

Query Match 1.0%; Score 49.6; DB 2; Length 1430;
Best Local Similarity 44.3%; Pred. No. 0.0031;
Matches 202; Conservative 0; Mismatches 254; Indels 0; Gaps 0;

QY 3085 ggcgaatataagatgcgaatcaagagtgacgtcgtcaagagcagcttgaagaaga 3144
DB 521 ggcgtactcgagatataacacacacacacacacacacacacacacacacacac 580
QY 3145 ctcaactcactgaacaatccgaaacacgtactcgaagacttctcagtgcttcacaag 3204
DB 581 AACACACACAAATGCTACTCCGAGAAATTACACACACAAACAAATGCTCTCAAGAAAT 640
QY 3205 aagaagaagccgagatgcgagagacacacacacacacacacacacacacacacac 3264
DB 641 TACACACACAAATGCTACTCCGAGAAATTACACACACAAACAAATGCTCTCAAGAAAT 700
QY 3265 aaac 3324
DB 701 AACACACACAAAGGCTTCTCCGAGAGTTACACACACAAACAAACAAATATTTTC 760
QY 3325 tccgagagagacatccagacagagatactacagcagctcgaagacttcaagtgctc 3384
DB 761 TCCGAGAAATTTACACAAACAAATATTTCTCCGAGAAATTTACACACAAAT 820
QY 3385 tccaagctcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaag 3444
DB 821 AATACACAAATATTTCTTCTCCGAGAAATTACACACAAATATGCTTCTCCGAG 880
QY 3445 ctctctagcagactcgtac 3504
DB 881 AACTACACACAAATATGTTTCTCCGAGAAATTACACACAAACAAATATGCT 940
QY 3505 tacaccgagcaatagcccaagcgaagataataacagac 3540
DB 941 TTCTCTGAGAAATTTACACAAACAAACAAACAAACAAATGCC 976

RESULT 12
US-08-276-452A-24
Sequence 24, Application US/08276452A
Patent No. 5646029

GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shao-Lim
APPLICANT: Du, He
APPLICANT: Gane, Alison M
APPLICANT: Bagic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabidopsis Protein (Agp) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado

COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,452A
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 49617824

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 1690 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 60..1442

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1..38

OTHER INFORMATION: /note- "Nucleotide sequence

OTHER INFORMATION: obtained by PCR which does not overlap with the

OTHER INFORMATION: cDNA clone"

FEATURE:

NAME/KEY: misc.feature

LOCATION: 60..128

OTHER INFORMATION: /note- "Predicted transmembrane

OTHER INFORMATION: segment"

FEATURE:

NAME/KEY: misc.feature

LOCATION: 135..179

OTHER INFORMATION: /note- "Derived amino acid sequence

OTHER INFORMATION: corresponding to the peptide sequence by protein

OTHER INFORMATION: microsequencing"

FEATURE:

NAME/KEY: misc.feature

LOCATION: 135..179

OTHER INFORMATION: /note- "Amino acids 27 to 36, 38,

OTHER INFORMATION: and 40 are identical to that in the peptide

OTHER INFORMATION: obtained by direct microsequencing"

FEATURE:

NAME/KEY: misc.feature

LOCATION: 135..179

OTHER INFORMATION: /note- "Amino acid 26 may also be

OTHER INFORMATION: Ala; 37 and 39 can also be undetermined residues"

US-08-276-452A-24

Query Match 1.0%; Score 48.2; DB 1; Length 1690;
Best Local Similarity 43.2%; Pred. No. 0.0081;
Matches 230; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

QY 859 gaccttcacatcatacaaacagcttgaagaagccataactcactcagcgtacgtgag 918
DB 573 GATGAGTCTTCAATTAACAAATGTTATGATTCACAAACACACACACAAATGTTAT 632
QY 919 aagcgatagacacccctcaagaagaaatgaatatcaagaagactcgtcgaagaatatt 978
DB 633 GATTCACAAATTAACAAACAAATGAGATGAGGCTTCCGAGAAATTAACAAACAAAT 692
QY 979 gaaattaagaatctccgcagccaaactctggaacaccccttaacacgcgtgtgacaag 1038

Db 693 GGCTACTCGAGAAATGCTAATAACAAAATAAATGCTACTCAGAAATTACAAAC 752
QY 1039 aacaagaagatagaggagcaagagaagaatcaagaatcgcgaacacattagtlc 1098
Db 753 AATAACAACAATGGCTACGCCCAAAATTACAAATGGCTACTCTCAAGTTACAAAC 812
QY 1099 aacatagattctcttactcagatcccttgagctgaggtactacttgagagagaat 1158
Db 813 AACATAATTTTACTCGGAGAAATTACAAACAACAACAACAATGTTTTCTCGAGAT 872
QY 1159 aagaatagacatcccgcaaaagtcgagaacaaggaatcaaccgaacctaatgatat 1218
Db 873 TCCAACAACAATGGCTACTCCAAAAGATCAACATATATGGCTACTCCAGAAATTACATG 932
QY 1219 cccaatggtgagcgtacccctctgtctataagatatcaaacagctctcaagagctc 1278
Db 933 AACACAACAATGGCTTCTCCGAGATTACAAACAACAACAACAACAACAACAAC 992
QY 1279 aatagctcggtgacttgatttaaccccttgattatagaaagaaaccttaagatatc 1338
Db 993 AACGTTTCTCTGAGAAATTACAAACAACAACAATTAATGTTTTCTCCGAGAAATTAC 1052
QY 1339 tacacagacaatgagaagaagtttatacgaagaatcaaggaagagatcaa 1391
Db 1053 AACACAACAATTAACAACAATGCTTTTACGAGAAATTACAAACAACAACAACA 1105

RESULT 13
US-08-798-744-24
Sequence 24, Application US/08798744
Patent No. 5830747
GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shiao-Lim
APPLICANT: Du, He
APPLICANT: Gane, Alison M
APPLICANT: Bacic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabidnolactan Protein (AGP) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,744
FILING DATE: 13-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,452
FILING DATE: 18-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1690 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 60..1442
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..38
OTHER INFORMATION: /note= "Nucleotide sequence obtained by PCR which does not overlap with the OTHER INFORMATION: cDNA clone"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 60..128
OTHER INFORMATION: /note= "Predicted transmembrane OTHER INFORMATION: segment"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 135..179
OTHER INFORMATION: /note= "Derived amino acid sequence OTHER INFORMATION: corresponding to the peptide sequence by protein OTHER INFORMATION: microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 135..179
OTHER INFORMATION: /note= "Amino acids 27 to 36, 38, OTHER INFORMATION: and 40 are identical to that in the peptide OTHER INFORMATION: obtained by direct microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 135..179
OTHER INFORMATION: /note= "Amino acid 26 may also be OTHER INFORMATION: Ala; 37 and 39 can also be undetermined residues"
US-08-798-744-24

Query Match 1.0%; Score 48.2; DB 2; Length 1690;
Best Local Similarity 43.2%; Pred. No. 0.0081;
Matches 230; Conservative 0; Mismatches 303; Indels 0; Gaps 0;
QY 859 gacctgtccatctataacaacagctgaagaagccataactatcagcgtactgag 918
Db 573 GATGAGTCTCTCAATACAAATGATTATGATTCACAAACAACGACAAACAATGTTAT 632
QY 919 aagcgatagacaccctcaagaagaatgaataatcaagaactctcgaagaatrat 978
Db 633 GATTCCAACAATTAACAACAATGATGATGCTTCTCCGAGAAATTACAAACAAT 692
QY 979 gaattaaagatccctcgccagcaactctgggaacacccctaaacgctctgagacaag 1038
Db 693 GGCTACTCGGAGAAATGCTAATAACAAAATTAACATGGCTACTCGAATAATTACAAAC 752
QY 1039 aacaagaagatagaggagcagcagaagaatcaagaagatcgcgaacacattagtlc 1098
Db 753 AATAACAACAATGGCTACGCCAAGAAATTACAAATGGCTACTCTCAAGTTACAAAC 812
QY 1099 aacatagattctcttactcagatcccttgagctgaggtactacttgagagagaat 1158
Db 813 AACATAATTTTACTCGGAGAAATTACAAACAACAACAACAACAATGTTTTCTCGAGAT 872
QY 1159 aagaatagacatcccgcaaaagtcgagaacaaggaatcaaccgaacctaatgatat 1218
Db 873 TCCAACAACAATGGCTACTCCAAAAGATCAACATATATGGCTACTCCAGAAATTACATG 932
QY 1219 cccaatggtgagcgtacccctctgtctataagatatcaaacagctctcaagagctc 1278
Db 933 AACACAACAATGGCTTCTCCGAGATTACAAACAACAACAACAACAACAACAAC 992
QY 1279 aatagctcggtgacttgatttaaccccttgattatagaaagaaaccttaagatatc 1338
Db 993 AACGTTTCTCTGAGAAATTACAAACAACAACAATTAATGTTTTCTCCGAGAAATTAC 1052
QY 1339 tacacagacaatgagaagaagtttatacgaagaatcaaggaagagatcaa 1391

Db 744 CGATTTAGAACAAGAGAGCGTCTTAAGAAAAAGTTGCAAGAACAAACGATTAG 803
Oy 869 tctatacaaacagcttgagaagcccatcaaccctcagcgtactggagaagcgcatag 928
Db 804 ACAAGAGACACTTGTCTAAAGAAAAAGTTACAGAGCGCAAGCGATTGTGACAAACATAG 863
Oy 929 acaccctcaagaagaatgaatatcaagaactgctgcgaagaatgaatga 980
Db 864 ACTTGCTTAAGAAAAAGTTGCAAGAACAAACGATTAGAACAGAGAGA 915

RESULT 18
US-08-098-327E-34
; Sequence 34, Application US/08098327E
; Patent No. 6270771

GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUILHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: MCGOWAN, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-34

Query Match 0.9%; Score 46.4; DB 4; Length 988;
Best Local Similarity 45.7%; Pred. No. 0.018;
Matches 161; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

Oy 629 tgaagtcagagccagcgttggaagctatgaagaagtgtgcttcgatatcgcaagc 688
Db 596 TAAAGAAAAGTTGCAAGAACAAACGATTGTGAAACAAGAGAGCGTCTTAAGAAAA 655
Oy 689 cctctacacacatcaagaagaatgttggaagaatgaagaatlatataaagaataaga 748
Db 656 GTTGCAAGAACAAACGATTGTGAAACAAGAGAGACTTGTCTTAAGAAAAAGTTGCAAGA 715

Oy 749 agaccatcgagaacatlaacgagctgatlgaagaatcccaaaaagaccatagaacaaata 808
Db 716 ACAACAAACCGATTGTGAAACAAGAGAGCGTCTTAAGAAAAAGTTGCAAGAACAAACG 775
Oy 809 agaatgcaccccaaggaagaagaagaagtgtgaccagcccgtagcccgctga 868
Db 776 CGATTTAGAACAGAGAGAGCGTCTTAAGAAAAAGTTGCAAGAACAAACGATTAG 835
Oy 869 tctatacaaacagcttgagaagcccatcaaccctcagcgtactggagaagcgcatag 928
Db 836 ACAAGAGACACTTGTCTAAAGAAAAAGTTACAGAGCGCAAGCGATTGTGACAAACATAG 895
Oy 929 acaccctcaagaagaatgaatatcaagaactgctgcgaagaatgaatga 980
Db 896 ACTTGCTTAAGAAAAAGTTGCAAGAACAAACGATTAGAACAGAGAGA 947

RESULT 19
US-08-462-625-34
; Sequence 34, Application US/08462625
; Patent No. 6319502

GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUILHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,625
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/098,327
FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: MCGOWAN, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-462-625-34

Query Match 0.9%; Score 46.4; DB 4; Length 988;
Best Local Similarity 45.7%; Pred. No. 0.018;
Matches 161; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

ADDRESS: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axameley Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DHS alpha

TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420

Query Match	0.98;	Score 45.6;	DB 6;	Length 697;
Best Local Similarity	47.08;	Pred. No. 0.023;		
Matches: 141;	Conservative	0;	Mismatches 159;	Indels 0;
			Gaps	0

[illegible]

```

RESULT 24
PCT-US95-10668-1
; Sequence 1, Application PC/TUS9510668
; GENERAL INFORMATION:
; APPLICANT: James Eberwine

```


Query Match	0.9%	Score	44.8	Length	950
Best Local Similarity	45.5%	Pred.	No. 0.046		
Matches	160	Conservative	0	Mismatches	192
				Indels	0
				Gaps	0

629 tgaagatcagagccaacagagtttgacgtatttgaagaagtgtcttcgatatcgcaagc 688

Db 29 TAAAGAAAGTTGCAAGACACAAAGGATTTAGACAGATAGACTTGGCTAAGAGAAA 88

[illegible]

09 G1ACAGGACAGCAGAGCGA11AGAGCACAGAGAGC11GC1AAAGAGAGAGAG11GCAGGA 140

UU

QY 749 agaccatcgagaacattaacgagctgtagcgaagaatccaaaagaaccatagacaaaara 808

Db 149 ACAACAAAGCGATCTAGAACAAAGAGAGCGTGCTAAAGAAAAGTTGCAGAACAACAAAG 208

Db 209 CGATTAGACACAGAGACGCTGCTAAGAAAGTTGCAAGACACACAAGCGATTTGA 268

QY 869 tctataacaacagcgttgaagaagccataacctcatcagcgttacttggagaagcgcatag 928

Db 269 ACAGATGACCTTCTAAAGAAAAGTTACAAGAGCAGCAAGCGATTAGAACACAGAG 328

929 acacccctcaagaaagatgaaatatcaaaagaaactgctccacaagaattatga 980

[illegible]

UD 329 ACGTCTATTGAGAAATGTTGCTAAGATCATCAATGCGATTTATGATCTATATATGTA 300

RESULT 28

US-08-462-625-32
: Sequence 32, Application US/08462625

Patent No. 6319502
GENERAL INFORMATION:

GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine

;; APPLICANT: DRULHE, Pierre
;; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE

; TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPIITOPES CAPA
 ; TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES

NUMBER OF SEQUENCES: 46

ADDRESS: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404
CITY: Alexandria

STATE: Virginia
COUNTRY: United States

ZIP: 22313-1404

```

;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk

```

```

;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA.

```

COMMENT REFLECTION DATA:
APPLICATION NUMBER: US/08/462,625
FILING DATE OF THIS 1005

FILED DATE: 05-JUN-1995
CLASSIFICATION: 424

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/098,327
 ;

FILING DATE: 24-NOV-1993
 DTOD ADDITION DATA.

PR 91 01286

ATTORNEY/AGENT INFORMATION:
FILING DATE: 05-FEB-1991

NAME: MCGOWAN, MALCOLM K.

RESULT 30
 US-08-098-327E-45
 Sequence 45, Application US/08098327E
 Patent No. 6270771
 GENERAL INFORMATION:
 APPLICANT: GUERIN-MARCHAND, Claudine
 APPLICANT: DRUILHE, Pierre
 TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
 TITLE OF INVENTION: HEPATIC STRAES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
 OF STIMULATING THE T LYMPHOCYTES
 NUMBER OF SEQUENCES: 46
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Burns, Doane, Swecker & Mathis
 STREET: P. O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 00:14:21 : Search time 5257.13 Seconds
(without alignments)
12682.768 Million cell updates/sec

Title: US-09-269-874a-2

Perfect score: 4940

Sequence: 1 cgcacgcagatgaataatcat.....ttcatcataatgacgatg 4940

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	235	4.8	537	10	BI815756 PESTOaa3
2	194	3.9	500	9	AU088128 AU088128
3	168.8	3.4	500	9	AU086832 AU086832
4	161.2	3.3	501	10	BI814631
5	158	3.2	500	9	AU087573 AU087573
6	156.4	3.2	500	9	AU088129 AU088129
7	134.4	2.7	303	10	BI815239 PESTOaa1
8	124.4	2.5	408	10	BI814931 PESTOaa0
9	119.8	2.4	365	10	N97742 N97742
10	112.2	2.3	587	10	BI815957 PESTOaa4
11	108.8	2.2	500	9	AU086246
12	108.8	2.2	755	10	BM159731 EST562254
13	107.6	2.2	753	10	BM160423 EST562246
14	105	2.1	678	10	BM165725 EST568248
15	102.6	2.1	402	10	BI670649 PESTOaa0
16	101.6	2.1	400	10	N97689
17	98.2	2.0	354	10	N97605 N97605

18	97.2	2.0	647	12	A2526918
19	91.2	1.8	660	10	BM159472
20	91.2	1.8	777	10	BM170928
21	89.6	1.8	732	10	BM160106
22	88.4	1.7	313	10	T18122
23	82.8	1.7	282	10	T18003
24	80.6	1.6	878	12	BM153470
25	79	1.6	483	10	BM165179
26	79	1.6	703	10	BM167609
27	79	1.6	769	10	BM16688
28	78.8	1.6	1147	12	BI13042
29	74.4	1.5	344	10	BM159366
30	73.8	1.5	540	12	A2813205
31	73.4	1.5	948	10	BM415494
32	71.6	1.4	1885	10	BE420745
33	69.8	1.4	691	12	BH040136
34	68.6	1.4	682	10	BM170771
35	68.6	1.4	908	12	A2548467
36	68.4	1.4	596	12	A2460541
37	68	1.4	907	12	CNS021174
38	67.2	1.4	500	9	AU087665
39	66.6	1.3	810	3	BI644882
40	66.4	1.3	827	12	CNS02156
41	65.8	1.3	731	12	A2191902
42	65.6	1.3	960	12	CNS005F3
43	65.4	1.3	783	12	CNS01112
44	65.4	1.3	897	12	CNS07AB2
45	65.4	1.3	1135	12	CNS03360

ALIGNMENTS

RESULT 1
BI815756
LOCUS
DEFINITION
BI815756 537 bp mRNA linear EST 03-OCT-2001
PESTOaa32c09.y1 Plasmodium falciptarum 3D7 asexual cDNA Plasmodium
falciptarum cDNA 5' similar to TR:Q9TVG8 Q9TVG8 MSAL PROTEIN ;, mRNA
sequence.
ACCESSION
BI815756
VERSION
BI815756.1 GI:15909334
KEYWORDS
EST
SOURCE
Malaria parasite P. falciptarum.
ORGANISM
Plasmodium falciptarum
REFERENCE
1 (bases 1 to 537)
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS
Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Matta,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagaris,I., Richey,J., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Richey,J., Madkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
TITLE
Washu Plasmodium EST Project
JOURNAL
Unpublished (2001)
COMMENT
Contact: L. David Sibley
Washu Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@orcim.wustl.edu), Washington University
Seq primer: -400p from Gibco
High quality sequence stop: 422.
Location/Qualifiers
1. 537
/organism="Plasmodium falciptarum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciptarum 3D7 asexual cDNA"
/lab_host="DH10B (Genehogg, Invitrogen, Inc.)"

FEATURES

source


```

QY 4834 tatcacctctgcagcgggatttttgcctcacgcttaattcttcgggacatccctccctc 4893
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 TATCCACTTTTCTGATTTTCTGCAGTTCCTTAACCTTTAGGAATATCATTTCTTA 262

QY 4894 ctgatacctatgctgatacctcgtacagctcatcattataagtcgaty 4939
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 261 TTATATCTCATGTTAATATTAATACAGTTTCATTTAAAAAATGTTAGG 216

RESULT 5
AU087573 500 bp mRNA linear EST 27-JAN-2001
LOCUS AU087573 Sugano Malaria cDNA library Plasmodium falciparum cDNA
DEFINITION clone XPFn5232 similar to P.falciparum gp190 (MSAL, MSP1, PMMSA)
          for precursor of major merozoite surface antigens, mRNA sequence.
ACCESSION AU087573
VERSION AU087573.1 GI:12389714
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 500)
AUTHORS Matanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE FULL-malaria: a database for a full-length enriched cDNA library
          from human malaria parasite, Plasmodium falciparum
JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE 20574754
COMMENT Contact: Junichi Matanabe
          Institute of Medical Science
          The University of Tokyo, Department of Parasitology
          4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
          Tel: 81-3-5449-5378
          Fax: 81-3-5449-5410
          Email: jwatanab@manage.ims.u-tokyo.ac.jp
          Suzuki,Y., Yoshitomo Nakagawa,K., Maruyama,A. and Sugano
          S. Construction and characterization of a full length-enriched and
          a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
source
1..500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 237 a 52 c 58 g 153 t
ORIGIN
Query Match 3.2%; Score 158; DB 9; Length 500;
Best Local Similarity 59.3%; Pred. No. 5.7e-21;
Matches 288; Conservative 0; Mismatches 195; Indels 3; Gaps 1;

QY 1533 ctccggagaatgatggaagaaacgctactcttcaaaagtggagaagaaatgcacacata 1592
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15 CGTAGATTAATAATTTTCAGTGCAGATATACATTAATGTTGAATAAACAAATTAATAA 74

QY 1593 taaccttcacatcagaagattcctaagcataactctgagaagctccacaagctcttaa 1652
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 TAAATTTTCTCTTAATATTTCTGTATTAATGTTCAAAAATTAAGAGCTCTTTC 134

QY 1653 gataatgagagactatctctcgcgaacatctgtgtgagaagaagaactaaagatacaaa 1712
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 ATATCTTGAAGATTATCTTTAAGAAAGAAAGAAATTTTGAAAAAGATTTTATCATCTATTA 194

QY 1713 gaatcctatagaataatcgaaaagatcgagagctgtgtgagaacttaagaaga 1772
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 TACTTTGAAAAATCGCCCTCAACCTGATTAATAAAAAAATTAACGAAGAAATTAAGACTAG 254

QY 1773 tgaagaacacgtgtgtgagaagaagattacaaaagaagaataaaccagatgagaagat 1832
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 TGAACAACAAATTTCTAGAAAAAATTTTAAAGAGCTAACACATTTCAGCAATG---GTTC 311

```

```

QY 1833 cctgagagctccgatatgtttaaagtcagagagtgctccatgaacaagat 1892
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 CTTAGAAGTATTCGATATTTCTAATAATTAACAGTACAAAAGTTTATTATTAATAAAAAAT 371

QY 1893 tgaatgaactaagaagacataactcattctgagaagcgtggagtttaaacataataca 1952
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 AGAAGACTTAGAAGAAAGATGAAATTAATTTTAAAAATGCAACACTAAAGATAGTATTC 431

QY 1953 tctgcgaatagttataagcagaagataagcagaacataactcactcgtactcaa 2012
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 432 TGTAACCAATATTTATTAACCAACAAATTAACCAACATATTATTATTAATGATTAATA 491

QY 2013 gaaaga 2018
      |||||
Db 492 AAAAGA 497

RESULT 6
AU088129 500 bp mRNA linear EST 27-JAN-2001
LOCUS AU088129 Sugano Malaria cDNA library Plasmodium falciparum cDNA
DEFINITION clone XPFn560 similar to P.falciparum (ME7) gene for merozoite
          surface antigen 1, mRNA sequence.
ACCESSION AU088129
VERSION AU088129.1 GI:12390270
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 500)
AUTHORS Matanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE FULL-malaria: a database for a full-length enriched cDNA library
          from human malaria parasite, Plasmodium falciparum
JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE 20574754
COMMENT Contact: Junichi Matanabe
          Institute of Medical Science
          The University of Tokyo, Department of Parasitology
          4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
          Tel: 81-3-5449-5378
          Fax: 81-3-5449-5410
          Email: jwatanab@manage.ims.u-tokyo.ac.jp
          Suzuki,Y., Yoshitomo Nakagawa,K., Maruyama,A. and Sugano
          S. Construction and characterization of a full length-enriched and
          a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
source
1..500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 197 a 65 c 58 g 144 t 36 others
ORIGIN
Query Match 3.2%; Score 156.4; DB 9; Length 500;
Best Local Similarity 57.6%; Pred. No. 1.2e-20;
Matches 250; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 3161 atccgaacacgactgacgaactctcagtgcttccaagaagaagaagccgaga 3220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 57 AACCCATATGATTAACAAACATTTCTGTTNCTTTAACAANTNAAGAAAGAGCTGAAA 116

QY 3221 tcgcgagaagaagaacacactcgtgagaacacaaagattcttcacaactataaagac 3280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 117 TACGAGAACTGTAAANNCATTAAAAACGACAGANTATTATTGAACATTATTAAGAGAC 176

QY 3281 tcgtcaagataataatgagagctctctcctctgaagactcctccgagaagaagatcc 3340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 177 TTGGTAATATTTATTAATGATGATCTTCATTAATAAACTTAAGTGAAGTATCAATTC 236

```


OY	3341	aagacgagatatactaacgcgaaccctcgagaacttcaaggctcgtctaagctcgaaagca	3400
Dd	237	AAACGAGAGATATTATATGCCAATTAGANNAANNNTAGACTATTAACTNNAATAAATGGAA	296
OY	3401	agctgaagjacacaacctgacccgtagaagaagaagctcacgtacctctctagcgaatgc	3460
Dd	297	AACTCGAATGATTATTATTACATTTAGGAGAAGAAAATTTATCTTCTTATCACAAGTGATTC	356
OY	3461	atccactgatcgccgagctccaaggaagtcatlaagaacaaagaatacaccggcaatagcc	3520
Dd	357	ATCATTTTATTTACTGTAAATTAAGAAAGATTATNAANAANTTAATTAATACAGTATTC	416
OY	3521	caagcgagaataatactcacgcgtgataaatacgcactggaattcttaagaagaatgccgctg	3580
Dd	417	CAGGAGANNAATTAAGAAAGTTAACGAACTTTNAAAATCTTACNAAAAAAATTTTTTNCCA	476
OY	3581	aagagacagatgctc	3594
Dd	477	NAAAGCAAAGTTNC	490
RESULT	7		
B1815239		303 bp	mRNA linear EST 03-OCT-2001
LOCUS		PfSTeoaal6bl.1.y1 Plasmodium falci-parum 3D7 asexual cDNA Plasmodium	
DEFINITION		falciparum cDNA 5' similar to TR:O90AI8 O90AI8 SURFACE PROTEIN-1 ;,	
ACCESSION		B1815239	
VERSION		B1815239.1	GI:15908372
KEYWORDS		EST.	
SOURCE		Malaria parasite P. falciparum.	
ORGANISM		Plasmodium falciparum	
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS		I (bases 1 to 303)	
		Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D.,	
		Marras, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,	
		Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,	
		Tsangarelsivilli, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,	
		Maguire, L., Ritchey, J., Wadkins, J., Kennedy, S., Levinso, D.,	
		Waterston, R., Wilson, R., and Sibley, D.	
TITLE		Washu Plasmodium ESP Project	
JOURNAL		Contact: L. David Sibley	
COMMENT		Unpublished (2001)	
		Washu Plasmodium ESP Project	
		Washington University School of Medicine	
		4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA	
		Tel: 314 286 1800	
		Fax: 314 286 1810	
		Email: estewartson.wustl.edu	
		Library was constructed by Debopam Chakrabarti DNA sequencing by:	
		Washington University Genome Sequencing Center for information on	
		obtaining a clone please contact: L. David Sibley	
		(sibley@borcim.wustl.edu), Washington University	
		Purative full length read	
		vector to vector length is 433	
		Seq primer: -40UP from Glibco	
		High quality sequence stop: 257.	
FEATURES		Location/Qualifiers	
source		1..303	
		/organism="Plasmodium falciparum"	
		/db_xref="taxon:5833"	
		/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"	
		/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"	
		/note="Vector: pluescript SK plus; Site 1: EcoRI; Site 2:	
		XhoI; Library was constructed by Debopam Chakrabarti.	
		Total RNA samples were isolated from mixed stage	
		sapron(0.1%) lysed P. falciparum 3D7 infected	
		erythrocytes by the acidic guanidinium-phenol chloroform	
		method. The poly A+ RNA was isolated by the poly(A)-tract	
		mRNA isolation system (Promega, WI) using streptavidin	
		Magneshere particles. Directional cDNA libraries were	
		constructed by oligo d(T) priming of poly(A)+ RNA (5mg)	
		into EcoRI and XhoI sites of 1 ZapII vector using the Zap	

BASE COUNT	106 a	41 c	41 g	115 t
ORIGIN				
Query Match	2.7%	Score 134.4	DB 10;	Length 303;
Best Local Similarity	72.5%	Pred. No. 2.4e-16;		
Matches 174;	Conservative	0;	Mismatches 66;	Indels 0;
			Gaps 0;	
Qy	4700	ataagtcgtggaagaaacccaacccctaccctgcaatgaaacaaatggcgggtgtgacgcg	4759	
Db	1	ATTAATGTGTTGAAATCCAAATCCTACTCTGTAACGAAAAATATAGTGTGATGTGCAG	60	
Qy	4760	atgctcaatgcaacggagaagacagcggctctaaaggaaagaaaatcatcgtcgagtta	4819	
Db	61	ATGCCACATGTATCCCAAGAGATTTCAGGTAGCAGCAGAAAAAATCAATCATGTGATTA	120	
Qy	4820	ctaaagccgactccctcatccatctctgcagcggattttgttcacagctcatatccctgg	4879	
Db	121	CTAAACCTGATTTCTTATTCACCTTTTCGATGGTATTTTTCGACATTCCTCTACTTTAG	180	
Qy	4880	gcatctctctcctgctgatacctcatcgtatcctgctgatacagctcatcataagatcga	4939	
Db	181	GAAATATCATTTCTTATTAATCACTGATTAATATTAATATTCAGTTTCATTTAAAAAATGTACG	240	
RESULT	8			
LOCUS	BiB14931	408 bp	mRNA	linear
DEFINITION	PFSToa08b05.y1 Plasmodium falci-parum 3D7 asexual cDNA Plasmodium falci-parum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ; , mRNA sequence.			
ACCESSION	BiB14931			
VERSION	BiB14931.1	GI:15907780		
KEYWORDS	EST.			
SOURCE		malaria parasite P. falci-parum.		
ORGANISM		Plasmodium falci-parum		
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS		1 (bases 1 to 408) Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dente, M., Theising, B., Bowers, Y., Gibbons, M., Ritten, E., Bennett, J., Jentes, E., Ronko, I., Tsagaris, I., R., Belagayrod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Watkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R., and Sibley, D.		
TITLE		Washu Plasmodium EST Project		
JOURNAL		Contact: L. David Sibley		
COMMENT		Washu Plasmodium EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu Library was constructed by Debopam Chakrabarti. DNA sequencing by: Washington University Genome Sequencing Center for Information on obtaining a clone please contact: L. David Sibley (sibley@orcim.wustl.edu), Washington University Purative full length read vector to vector length is 409 Seq primer: -40UP from Gibco.		
FEATURES	source	Location/Qualifiers		
	1..408	/organism="Plasmodium falci-parum"		
		/db_xref="taxon:5833"		
		/clone_lib="Plasmodium falci-parum 3D7 asexual cDNA"		
		/lab_host="DH10B (Genesys, Invitrogen, Inc.)"		
		/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:		


```

/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db.xref="taxon:73239"
/clone="PYCUG14"
/clone_1lb="PYBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/notes="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZap arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybridZap vector and plasmid DNA
isolated."

```

```

BASE COUNT      268 a      76 c      92 g      161 t
ORIGIN
Query Match      2.3%; Score 112.2; DB 10; Length 597;
Best Local Similarity 52.4%; Pred. No. 7e-12;
Matches 292; Conservative 0; Mismatches 238; Indels 27; Gaps 1;

```

```

Oy 487 aaccatcagcttcaaatatctgttgcgttcaagagatcaatgaactcctgtac 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36 AATGATATGGAGTAAATATTATTGAAAGTAAGAAGATTAAATGATTAATGCAC 95
Oy 547 aagttgaattctactcgtactgtcgaagggccaactgaatgacgttgcgcgaatgac 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 GCATTAATTTTATTATGATGTAAGTAAATTAATGATATGTCGCAATATAT 155
Oy 607 tatgtcaaatcattcaatttgaagatcagagccaagsgtctgacgtattgagaag 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 TATGTGAAATTCCTGAAACATCTTAAATTTAGTGAAGAAGAACAGAAATGCTTAAAAA 215
Oy 667 ttgtcttcgcatatcgaagcctcgcacacatcaagcaatgctgtggaagtgtaa 726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216 GTAAATTTAGGTTATGAAACCAATAGAAATATTCAAGACGATTTGAAAGTTAAGA 275
Oy 727 gattatataaagaataaagaagacacacgaacattcaagcgtgatcgaagaatccc 786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 ATTTACATAGAAAGAAATTAAGAAGAACTGTCAGCTTTAAAGCGCTTATTTGCTGAAGA 335
Oy 787 aaaaagaccata-----gacaaaataagaatgcaacc 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 ACAAAAAAATATACAACTGAAGGTAAACGAGATTCGAATGCGTGTGATAGCGCAT 395
Oy 820 aaggaagagaagaagaagaagttgtaccagagccagtagcgccttccatcattacaac 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 AATAATTAATAAAAAAACAATATACCAAGCTATGTCAATGTATTTTATACAAAAA 455
Oy 880 cagcttgaagaagaccataaccatcagcgtactctggaagcgcgatagacacccccaag 939
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 CAATTAGCTGAATATCAAAAGGTTGTGAGAGCTTGAAGAAAAAGGTTTCTACATTAAG 515
Oy 940 aagaatgaataatcaagaactcgtcgaagaagalttaagaaatcaatccctccgca 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 516 AAAATGATGCCATCAACCAATTTAGCGACAAATTTGAAGTTTCAATAGCTGCCCCGCTC 575
Oy 1000 gccaaactctgggaacac 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 GTCACTGCCGAACAC 592

```

```

RESULT 11
LOCUS      AU086246
DEFINITION AU086246 Sugano Malaria cDNA library Plasmodium falciparum cDNA
            clone XPhn2175 similar to P.falciparum gp190 (MSA1, MSP1, PMMSA)
            for precursor of merozoite surface antigens, mRNA sequence.
ACCESSION  AU086246
VERSION     AU086246.1 GI:12388387
KEYWORDS   EST.
SOURCE      malaria parasite P. falciparum.
ORGANISM   Plasmodium falciparum
REFERENCE  1 (bases 1 to 500)
AUTHORS   Matanabe,J., Sasaki,M., Suzuki,Y., and Sugano,S.
TITLE      WU:malaria: a database for a full-length enriched cDNA library
           from human malaria parasite, Plasmodium falciparum
JOURNAL    Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE    20574754
COMMENT    Contact: Junichi Matanabe
           Institute of Medical Science
           The University of Tokyo, Department of Parasitology
           4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
           Tel: 81-3-5449-5378
           Fax: 81-3-5449-5410
           Email: jwatanabem@nag.ims.u-tokyo.ac.jp
           Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
           S. Construction and characterization of a full length-length-enriched and
           a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
source
Location/Qualifiers
1..500
/organism="Plasmodium falciparum"
/strain="3D7"
/db.xref="taxon:5833"
/clone="XPhn2175"
/clone_1lb="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"

```

```

BASE COUNT      244 a      69 c      72 g      115 t
ORIGIN
Query Match      2.2%; Score 108.8; DB 9; Length 500;
Best Local Similarity 58.2%; Pred. No. 3.2e-11;
Matches 191; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```

```

Oy 1745 agacgctgttgagacacttaagaagatgtaagacagttgttggagaagaattaca 1804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19 ATACATATTAACGAAGAATAAAGAGTGTGAAACAAATTTCTAGAAAAAATTTA 78
Oy 1805 aagaagaaataaacaagatgagaagatccctggaggtccgatatgtttaagtccaag 1864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 AAGGACTAACACATTCCACCAATGCTTCCTTAGAAGTATCTGATATTTGTAATTAACAAG 138
Oy 1865 tgcagaaggtgcctccatcgaagaagattgaaactcaagaagactcaatccttga 1924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 TACAAAAAGTTTATTATTAAAAAATAAGAACTTTAAGAAAGATAGAAATTAATTTTAA 198
Oy 1925 agaagctgagtttaaaatataatctacatgtgcgaataagttataagcagagaataagc 1984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 AAAATGCAACAACCTTAATAAATAGTATTCATGTACCAAAATTTATTAACACCAAAATTAAC 258
Oy 1985 aggaaccaatacctacatctctactcaagaagaagatagacaactgaagtgatcagc 2044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 259 CAGAACCATATTATTATTATTATTAATAAAGAAAGATGATTAATTAAGATTTTATAC 318
Oy 2045 ccaaaagtcgagaagcctgatacaagaaga 2072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 CAAAAGTAAAGACATGTTAAAGAAAGA 346

```

```

RESULT 12
LOCUS      BM159731
DEFINITION EST562254 PYBS Plasmodium yoelii yoelii cDNA clone PYCUH95 5' end,

```

mRNA sequence.
 accession BM159731
 version BM159731.1 GI:17305412
 keywords EST.
 source Plasmodium yoelii yoelii.
 organism Plasmodium yoelii yoelii.
 reference Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 755)
 authors Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 title Plasmodium yoelii EST project at TIGR
 journal Unpublished (2001)
 comment Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 seq primer: ADF.

FEATURES
 source Location/Qualifiers
 1..755
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCJH95"
 /clone_1lb="PyBS"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (PAD-GAL4) was
 excised from the HybridZAP vector and plasmid DNA
 isolated."

BASE COUNT 323 a 103 c 136 g 193 t
 ORIGIN

Query Match 2.2%; Score 108.8; DB 10; Length 755;
 Best Local Similarity 52.4%; Pred. No.3.5e-11;
 Matches 285; Conservative 0; Mismatches 232; Indels 27; Gaps 1;

QY 487 aacattcatggtcacaatactgattgacggttacgaagaagatcaatgactcgtgac 546
 11 1111 111111 11111 11 11111 11111 11 1111
 Db 39 AATAGATGATGAGCTAAATATTGTAAGAAAGTAAGACAAATTTATGATTATTCAC 98
 QY 547 aagttgaattctactctgactgtctaaggccaactgaatgacgtttgcgcaatgac 606
 1 11111 11 11 11111 11 11111 11 11111 11 1111
 Db 99 GCAATAAATTTTATGATGATGACTTACGATTAATAATGATGATGCTGCAAAATAT 158
 QY 607 tatgtcgaattccattcaattggaatcagaagccaacaggttgacgcatgaagaag 666
 111111 11111111 11 11111 11 11111 11 11111
 Db 159 TATTTGGAATTCCTGGAACATCTTAAATATGAGAGAAACAGAAATGCTTAAAAA 218
 QY 667 ttggtcttcgagatcgaagcctctcgacaacatcaagaatggtggaaagatgaa 726
 1 11111 11111 11 11 11111 11 11111 11 11111 11111
 Db 219 GTAAATTTAGGTTATGAAAAACCAATAGAAATATTTCAAGCATATTTGAAAGTTGAA 278

QY 727 gattatattaagaagaataagaagaccatcgagaacatcaacgagctgacgaagatcc 786
 111 11 11111 11 1111 11111 1111 1111
 Db 279 ATTACATAGTAAGAAATTAAGAAACGTTCAGCTTTAAACGCTCTTATTCGTGAAGA 338
 QY 787 aaaaagaccata-----gacaaaataagaatgaacc 819
 1 1111 1111 11111 1111 1111
 Db 339 ACNAAAAAATFACMACCTGAAGTACGAAAGATTCGAATGACCTGTGTGTATACGAT 398
 QY 820 aagagagaagaagaagaagatgtgtacccagccagctacgacctgtccatatacaaa 879
 1 1 11111 11111 11111 1111 11111 11111
 Db 399 AATATATATATAAAAAAACCAATATACCAAGCATGTACATGTATTTTACAAAAA 458
 QY 880 cagcttgaagaagccataacctcatcagcgtactgagagaagcgcatagacacctcaag 939
 11 1111 1111 11 1111 11 1111 11 1111
 Db 459 CAATTAGCTGAAATFACMAAAGGTTGTGAAAGAAAGAGATTCTCATATTAAAG 518
 QY 940 aagaatgaatatcagaagacgtccgcgaagatgaatgaatgaatgaatcctcgcca 999
 111111 111111 11111 11111 11111 11111
 Db 519 AAAAATGATGCCATCAACCATTTATGGCAACAATTAAGATTCTCAATGCTGCCCGTC 578
 QY 1000 gccca 1003
 1 11
 Db 579 GTCA 582

RESULT 13
 BM160423 753 bp mRNA linear EST 04-DEC-2001
 LOCUS EST562946 PYBS Plasmodium yoelii yoelii cDNA clone PYCJH11 5' end,
 DEFINITION mRNA sequence.
 accession BM160423 GI:17306104
 version BM160423
 keywords EST.
 source Plasmodium yoelii yoelii.
 organism Plasmodium yoelii yoelii.
 reference Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 753)
 authors Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 title Plasmodium yoelii EST project at TIGR
 journal Unpublished (2001)
 comment Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 seq primer: ADF.

FEATURES
 source Location/Qualifiers
 1..753
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCJH11"
 /clone_1lb="PyBS"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column."

Size-fractionated cDNA was precipitated and ligated to HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated.

BASE COUNT 335 a 101 c 113 g 204 t
ORIGIN

Query Match 2.2%; Score 107.6; DB 10; Length 753;
Best Local Similarity 58.9%; Pred. No. 6e-11;
Matches 185; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Oy 487 aaccttcagctccaatcctgattcagcgttacgagatcaactcctgac 546
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 AATGATATGAGTAATAATTATTATTAAGTAAGAAATTAATGATTAATGCA 231
Oy 547 aagttgaattctactcgtactgctaaggccaactgaatgacgttcgcgaatgac 606
| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 GCATTAATTTTATATGATGCTTATAGATTAATTAATGATGCTGCAATATAT 291
Oy 607 tatgtcaaatccatcatttgaagatcagaccagcgttgacgtattgaagaag 666
||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 TATGTGAAATCTCGAACATCTTAATAATGTGAAGAAAGAAATCTTAAATAA 351
Oy 667 ttgcttcgatatcgcaagcctctgcacacatcaagcaatgttggaagaatgaa 726
| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 GTATTTTATAGTTATAGAAACCAATAGAAATATTCAAGCAGATTTGAAAAGTTAGAA 411
Oy 727 gattatataaagaagaagaagaccatcgagaacattaaagcgtgacgaagaatcc 786
||| ||| ||| ||||| ||| ||| ||| ||||| ||| ||| ||| |||
Db 412 ATTACATAGAAAGAAATAAGAAACGTGTGACAGCTTAAACGCTTATTGTCTGCAAGAA 471
Oy 787 aaaaagaccataga 800
| ||||| |||||
Db 472 ACAAATAAATACA 485

RESULT 14

BM165725

LOCUS 678 bp mRNA linear EST 04-DEC-2001
DEFINITION EST568248 PYBS Plasmodium yoelii yoelii cDNA clone PYCMR78 5' end,
mRNA sequence.

BM165725

BM165725.1 GI:17298957

EST.

KEYWORDS

SOURCE

ORGANISM

Plasmodium yoelii yoelii.
Plasmodium yoelii yoelii
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 678)

AUTHORS

Fraser, C.M. and Carucci, D.J.
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
Plasmodium yoelii EST project at TIGR

TITLE

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlone@tigr.org

JOURNAL

For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mri4.org/mr4pages/index.html

COMMENT

Seq primer: ADP.

Location/Qualifiers

1. 678

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCMR78"

/clone_id="PYBS"

/dev_stage="asexual blood stages"

FEATURES

source

/lab_host="E. coli XL-1 Blue"
/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

BASE COUNT 291 a 93 c 120 g 174 t
ORIGIN

Query Match 2.1%; Score 105; DB 10; Length 678;
Best Local Similarity 52.4%; Pred. No. 1.9e-10;
Matches 277; Conservative 0; Mismatches 225; Indels 27; Gaps 1;

Oy 502 aaatattcgtatgacgttacgaagatcaatgaaccctcgtaacgttaatttcac 561
||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 AAATTTTAAATTTGAAGAAAGAAATTTAATGATTAATGCAATTAATTTTAT 61
Oy 562 ttgcagctcgaaggccaactgacgttttcgccaatgacatgttcaattcca 621
| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 TATGATGCTTAAAGTAATAATTAATGATATGCTGCAATTAATTAATGCAATTTCT 121
Oy 622 ttcaatttgaagatcagaagccaagagttgacgtattgaagaagttgcttcgatat 681
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 GAACATCTTAAATTAATGTAAGAAAGAAACAGAAATGCTTAAATAAGTAAATTTAGTTAT 181
Oy 682 cgaagcctctgcgaacaatcaagaagacatgttggaagaatgagatattataaaga 741
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 AGAAACCAATATGAAATAATTTCAAGCAGATTTGAAAAGTTAATTAATTAATGCAAGA 241
Oy 742 aataagaagaccatcgagaacattaaagcgtgacgaagaatccaaagaacata--- 798
||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 242 AATTAAGAAACGTGTGACGCTTAAACGCTCTTATTCGTGAAGAAACAAAATAATACAA 301
Oy 799 -----gacaaaaataagaatgcacccaaggaggaagaag 834
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 302 CCGAAGGTAAAGCAATGCAATGACGCTGATGTCATACGATTAATTAATTAATAA 361
Oy 835 aagaagtgtaccagcgccagtaagactgtccatctatacaaacagcttgaagaagc 894
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 362 AAACCAATATACCAAGCTATGTACATGTATATTTCACAAAACAAATTTAGCTGAATA 421
Oy 895 cataaccatcagcgtactgagaagcgcatagacaccccaagaagaatgaataatc 954
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 422 CAAGAAGTGTGCGAAGTGTGAGAAAACGAGTTTCTACATTAAGAAATAAGATGCCATC 481
Oy 955 aaagaactgtcgcgaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1003
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 482 AAACCATTAATGCAACAAATTAAGTTCTCAATGCTGCCGCCGTCGTCGA 530

RESULT 15

B1670649

LOCUS 402 bp mRNA linear EST 12-SEP-2001
DEFINITION PEST0a02e11.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
falciparum cDNA 5' similar to FR:043995 043995 MEROZOITE SURFACE
PROTEIN-1, mRNA sequence.

B1670649

ACCESSION B1670649
VERSION B1670649.1 GI:15586033
KEYWORDS EST.

source malaria parasite P. falciparum.

	Tl8122	313 bp	mRNA	linear	EST 30-AUG-1994
LOCUS	Tl8122				
DEFINITION	0676c3 czapppdd2.1, Debopam Chakrabarti Plasmodium falciparum cDNA clone_0676c 5', mRNA sequence.				
ACCESSION	Tl8122				
VERSION	Tl8122.1	G1:462908			
KEYWORDS	EST,				
SOURCE	Malaria parasite P. falci-parum.				
ORGANISM	Plasmodium falciparum				
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
AUTHORS	1 (bases 1 to 313) Chakrabarti,D., Reddy,G.R., Dame,J.B., Almira,E.C., Laipis,P.J., Fenl,R.J., Yang,T.P., Rowe,T.C. and Schuster,S.M. Analysis of Expressed Sequence Tags from Plasmodium Falciparum Mol. Biochem. Parasitol. 66, 97-104 (1994)				
TITLE	95075403				
JOURNAL	Contact: Debopam Chakrabarti				
MEDLINE	Department of Molecular Biology and Microbiology				
COMMENT	University of Central Florida Orlando, FL 32816-2360 Tel: 407 384 2061 Fax: 407 384 3095 Email: dcchakepeganus.cc.ucf.edu				
	Seq primer: T3				
FEATURES	Location/Qualifiers				
Source	1..313				
	/organism="Plasmodium falciparum"				
	/strain="Dd2"				
	/db_xref="taxon:5833"				
	/clone="0676c"				
	/clone_lib="czappdd2.1, Debopam Chakrabarti."				
	/lab_host="E. coli XL-I blue"				
	/note="vector: Lambda ZAP II; Site_1: EcoR I; Site_2: Xho I; PolyA+ RNA, from asynchronous blood stage parasites of the Dd2 isolate cultured in vitro was reverse transcribed using an oligo dt-Xho I primer. Second strand was prepared using RNase H and DNA polymerase I. EcoR I adapters were ligated to the cDNA, and it was digested with Xho I. Prepared fragments were ligated into EcoR I + Xho I digested lambda ZAP II vector." "				
DNA COUNT	144 a 35 c 35 g 98 t 1 others				
ORIGIN					

	Query Match Similarity	1.8%; Score 88.4; DB 107; Length 313;
	Best Local Similarity	57.1%; Pred. No.3-2e-07;
Matches	180; Conservative	0; Mismatches 133; Indels 3; Gaps 1;
OY	1574	aagaactgcacacaaccataatccttgcacctgcatagtagaatctcaagcatacttgga 1633 Db 1 AAAAACAAAGATATAATATAAATTTCACCTTAATATTTGTGATATATGGTCAA 60
OY	1634	agctcaccaaagcctctaagtatatcgaggagactatttcctgcggaaacattgtttggaga 1693 Db 61 AATTAAAAAGGCCTTTTCATCATTCTTGAAAAGATTATCTTTAAGAAAGAATTTCTGA 120
OY	1694	aagaactaagratfaccagaacatccatcaagtaagaatcgaaaacgagatcgagacgttg 1753 Db 121 AAGATTTAATCATTTATTTACTTTGAAAACCTGGCCCTCAACCCTATATTAATAAATTTA 180
OY	1754	ttagaacatlaagaaggatgaagacaglttgtttgagaagaagaatlcaaaaaacgaaa 1813 Db 181 CAGAAGAATAAAGAGTAGTGTAACAACAAATTTCTAGAAAAAAATTTTTAAGACGATCAC 240
OY	1814	ataaaccsagatgagaagatcctlgsgagtlcccgatattgtttaagttccaagltgcgaag 1873 Db 241 ATTCCAGCAAAATG---CTTCCCTTAGAAGATGTCTGATATTTGTAATAATTACAGTACAAAAG 297
OY	1874	tgctcctcatgacaa 1888
Db	298	TTTYTAATATTTANA 312

REFERENCE	LOCUS	RESULT
1 (bases 1 to 282)	TI18003	23
Chakrabarti, D., Reddy, G. R., Dame, J. B., Almira, E. C., Laipis, P. J., Felt, R. J., Yang, T. P., Rowe, T. C. and Schuster, S. M.	TI18003	
Analysis of Expressed Sequence Tags from <i>Plasmodium falciparum</i>	0355c3 cbspfIB3.1, Debopam Chakrabarti <i>Plasmodium falciparum</i> cDNA	
Mol. Biochem. Parasitol. 66, 97-104 (1994)	clone 0325c 5', mRNA sequence.	
95075403	TI18003	
Contact: Debopam Chakrabarti	TI18003.1	
Department of Molecular Biology and Microbiology	GI:462789	
University of Central Florida	EST.	
Orlando, FL 32816-2360	Malaria parasite <i>P. falciparum</i> .	
Tel: 407 384 2061	<i>Plasmodium falciparum</i>	
Fax: 407 384 3095	Eukaryota; Alveolata; Apicomplexa; Haemosporida; <i>Plasmodium</i> .	
Email: dchakepegasus.cc.ucf.edu		
Seq primer: T3.		

BASE COUNT	ORIGIN	FEATURES
87 a	48 c	58 g
85 t	4 others	
		Location/Qualifiers
		1. .282
		/organism="Plasmodium falciparum"
		/db_xref="taxon:5833"
		/clone="0325c"
		/clone_lib="cdsPfHB3.1, Dehopam Chakrabarti"
		/lab_host="E. coli XL1-Blue"
		/note="Vector: pBluescript SK(+); PolyA+ RNA, from asynchronous blood stage parasites of the cloned Honduran HB3 isolate cultured in vitro, was reverse transcribed using an oligo dt-Xho I primer. Second strand was prepared using RNase H and DNA polymerase I. EcoR I adaptors were ligated to the cDNA, and it was cleaved with Xho I. Prepared fragments were ligated into EcoR I + Xho I double-digested pBluescript SK(+), and transformed E. coli XL1-Blue."

	Query Match	Similarity	1.7%	Score 82.8	DB 10	Length 282	
	Best Local	Similarity	62.0%	Pred. No.	4e-06		
	Matches	183	Conservative	0	Mismatches	75	Indels 37; Gaps 2
OY	143	tgttcacagaagagaagatggtgctgtaatgaaggacgagtgycacgycggttacaacca	202				
Db	4	TATTTCAAAAGGCAAAAATGCTTTTAATTAAGAGAAACAAGTCGAACAGCTGTGTTAACACTA	63				
OY	203	gcacacccggttctaaagggtcttggtctagcagcgtggtccggtgagtgcttggtgaccttg	262				
Db	64	GTACACCTGTGTTCAAAAGGTTTCAGTTGCTTCAGTGCGTTCAAGTGGCTCAGTTGGT---	119				
OY	263	ggggttcctgcgtccctccgycgycagcgtgycatcaagtggtctcaatgycgaacgycggtt	322				
Db	120	-----TCAGGTGGCTCAATTTGCTTCAGGTGGTT	147				
OY	323	ccgggaacagtcgaagaaccaatccatctgcaaacctctagcgatctcgacgcgaagtccct	382				
Db	148	CAGGTAATTTCAACAGCTGCACAAATCCTTCAGGTAATTTCAAGTATTTCAAGTGTAAATCTT	207				
OY	383	acgcgcacctcaagcaccgcagtgagaaacta-tctctcactatacaagaagctga	436				
Db	208	ACGGGATTTTAAACACAGAGTACGACGTAATTTACTTTGTTAACTATCAATAAGANCTCA	262				

RESULT	24		
BH153470			
LOCUS	878 bp	DNA	linear
BH153470			GSS 24-SEP-2001

QY	2965	aaacgtaagaagctcgggaaggagccatcaataagctgaagaagacactgtgaactgagctc	3024
QY	72	AATCAAAACAGAAGAACCAACCCACACAGAGAACTACACAGATCAAAACAAGAAGAC	131
Db	3025	gacctgtacacaagltacaaactgaacttggagaactcttcgacaagaagaagacatc	3084
QY	132	AACCACACCAAGAAAGAACCAACCAACAGAAAGAACCAACCAACACAGAAACAACAC	191
Db	3085	ggcagatataagatgtcagatccaagaagttgacctctgtcctaagagaggaacttgaagacaa	3144
QY	192	AACACAGGAAGAACCAACCAACAAAGAAAGAACACCAACAAAGAAAGAACCAACCAACAG	251
Db	3145	ctcacaccacttgaacaacttcogaaacacgttactcgaagacttcagatcttcagtgtcttccacaag	3204
QY	252	AAGAACCAACCAACCAAGAAAGAACAC-----CACACACAGAAAGAACCAACCAACAG	305
Db	3205	aagaagaagacccgagatcgcgcggagacagagaacactcttgagaaacccaagttcttc	3266
QY	306	AAGAACCAACCAACCAAGAAAGAACCAACCAAGAAAGAACCAACCAACCAACAGAAAGAC	365

Query ma
Best Loc

BASE COUNT	124 a	68 c	68 g	223 t
ORIGIN				

FEATURES	source
Db	1850 tctgttaagctccaagtgcagaaggtgcctccatcatgaagaattgatgaaccccaagaaga 1909
Db	321 TTGTAGAAATTGCGAATACAAAAAATCTTTATTATACCAAAACAAATTATGCAATTTAAATFMAA 262
Qy	1910 ctcaactctctctgaagaagcttgagatlaaacaataatacatatgctgcgaatgttata 1969
Db	261 CTGAAGATATTCTTTAAACAAGAGCTCAATTTAAAGCAAACTATATGTCTCCAAAAACATAG 202
Qy	1970 agcagagagaataagcaggaaccatactctcctcatctgctactcaagaagaagatagaacaac 2029
Db	201 GTAATGAAGGAGAAACCAACACCATCTACTTTTATATAGCTGTAAAAAAGAAAGTTGACAGAC 142
Qy	2030 tgaagaatgcatatgcccacaagctcgagagccctatcaacagaagaagaagaacttaaa 2088
Db	141 TTGCCCATTTATTCTCAAAAAATCGAAAGATGATGATTGCTTAAGAGAGAAAGAAATGGA 83
RESULT 26	
LOCUS	BM167609 703 bp mRNA linear EST 04-DEC-2000
DEFINITION	EST570132 PYBS Plasmodium yoelii yoelii cDNA clone PYCON62 5' end,
ACCESSION	BM167609
VERSION	BM167609.1 GI:17300841
KEYWORDS	EST.
SOURCE	Plasmodium yoelii yoelii.
ORGANISM	Plasmodium yoelii yoelii.
REFERENCE	Eukaryote: Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
AUTHORS	1 (bases 1 to 703) Carlsen, J.M., Baily, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B., Fraser, C.M. and Carucci, D.J.
TITLE	Plasmodium yoelii EST project at TIGR
JOURNAL	Unpublished (2001)
COMMENT	Contact: Jane Carlton Parasite Genomics Group The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301-530-9319 Fax: 301-838-0208 Email: carlone@tigr.org For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC http://www.malaria.mr4.org/mr4pages/index.html Seq primer: ADF.
FEATURES	Location/Qualifiers
source	1..703
	/organism="Plasmodium yoelii yoelii"
	/strain="17XL"
	/db_xref="taxon:73239"
	/clone="PYCON62"
	/clone_lib="PYBS"
	/dev_stage="asexual blood stages"
	/lab_host="E. coli XL-1 Blue"
	/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with PY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dt)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to Hybridzap arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the Hybridzap vector and plasmid DNA isolated."
BASE COUNT	294 a 113 c 125 g 171 t
ORIGIN	

```

Query Match      1.6%: Score 79; DB 10; Length 703;
Best Local Similarity 50.7%: Pred. No. 2.7e-05;
Matches 239; Conservative 0; Mismatches 205; Indels 27; Gaps 1;

QY 560 acttcgactctcctaaggcccaactcgaatlgacgttttcgcccattgactatctgtcacaatc 619
Db 1 ATTTATGATGTACTTATGAGATATAATTAAATGATATGTGTGCAAAATAATTATGTGAAATTC 60

QY 620 catlcaatttgaagatcagagccaaacgagcttggacgctatltgaagaagtgtgcttcggat 679
Db 61 CTGAACATCTTAAATAATTTGTTGTAAGAAACAGAAATGCTTTAAAAAGTAATTTAGCTT 120

QY 680 atcgcgaagcccttcgcacacatcaagsgcaatgttggaaagaatggaaatcattataa 739
Db 121 ATAGAAACCAATGAAATAATTTCAAGACGATATGAAAAGTTTACAAATTTTACATAGAAA 180

QY 740 agaataagaagacatcgtagaacattaacgagctgatcgaagaatccaaaaagaccata- 798
Db 181 GAAATTAAGAAACGTGTTGCAGCTTTAAACGCTCTTATTGTCTTAAGAAAACAAAAAATTC 240

QY 799 -----gacaaataaagaatcacaaccaagsgagaaga 832
Db 241 AACCGAAGTGTAACCAATTCGATGACATGACGCTAGTTGTGATGTCGATATAATTAATAAAA 300

QY 833 agaagaagctgtaccagggcccgtaacgagctgtccatctataacaacagcttgaagaag 892
Db 301 AAAAACCAATATACCAACCAATGTATGTAATTTTACAAAAACCAATTAAGCTGAAA 360

QY 893 cccataacctatcagctgacttggagaagcgcatatgacacccctcagaagaatgaaata 952
Db 361 TACAAAGAGTGTGCGAAGTCTTAGAAAAACGAGTTTCATTAATAAGAAAAATGATGCCA 420

QY 953 tcaagaagctcgtcgacaagattaaatgaatgaatccctcgcagacca 1003
Db 421 TCAAACCATTTATGGCAACAAATTGAAGTTCCTCAATGCTGCCCCCTGCTCA 471

RESULT 27
BM166688 mRNA EST 04-DEC-2001
LOCUS BM166688 769 bp linear
DEFINITION EST569211 PYBS Plasmodium yoelii yoelii cDNA clone pycnv80 5' end,
mRNA sequence.
ACCESSION BM166688
VERSION BM166688
KEYWORDS GI:17299920
SOURCE EST.
ORGANISM Plasmodium yoelii yoelii.
REFERENCE Plasmodium yoelii yoelii.
AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 769)
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
Fraser, C.M. and Garucci, D.J.
Plasmodium yoelii EST project at TIGR
Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ANCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES
source Location/Qualifiers
1..769
/oranism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="pycnv80"
/clone_id="PYBS"

```

```

/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/notes="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/CByu mice infected with Pyl/XTL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was washed
with XhoI and separated on a Sephadryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

```

```

BASE COUNT      340 a      101 c      104 g      224 t
ORIGIN
Query Match      1.6%; Score 79; DB 10; Length 769;
Best Local Similarity 46.6%; Pred. No. 2.8e-05;
Matches 285; Conservative 0; Mismatches 325; Indels 1; Gaps 1;

```

```

OY 2908 ttcaacttcgtgaatctaagccgatgatcaactctcttaacgatgaatctaa 2967
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111 TTCAATAATTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 170

OY 2968 cgtagaagcctggaagagacataagcgaagagacacactgctgac 3027
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 171 GTTAATCAATTAATCTTGAATCCCTCACTTAAGGAAATTAATCAAGAACATTAATATGAT 230

OY 3028 ctgtacaacaagtacaaactggaagagacactcttcgacaagaagaagcagtcg 3087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 221 CGTATCTACATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 290

OY 3088 aagttagaagcagatcaagaagtgactgctcgaagagcagccttgaaagcaactc 3147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 291 CTAACTCAATCGCAAAATTAAGATCTTAGTATATTGAAGACGATTTATTAATAAAGAAAA 350

OY 3148 aactcaactgaacaatccgaacaacgactcgaactctcagtgcttcttcaacaagaag 3207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 351 CAACTCTTAATGGCGTATTTTATATTAATGTTATGTAATTTCTTTAAACAAGAA 410

OY 3208 aaggaagcagatcgccgagacagagaacactctggaagacacagaattcttcaaa 3267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 411 AGAGAACTGAAAAAACAATATGTAGATTAATCAATTAATAAATAAATAAATAAATAA 470

OY 3268 cactcaagaagcctgctcaagatataatgcgagcttctcctctgaagcctctcc 3327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 471 TACTTCAAAAGCTCGTACATAAATTTTACTTCTGAAAGCTGTTTCTTTAAACAATTAATCT 530

OY 3328 gagagagcagatccagcagagataactacgacccctcgaagcctcgaagtcctgct 3387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 531 AAAGATACACTTGACAGAGAAATCCAAATTAATTTGAAAAATTCGAGACATACACT 590

OY 3388 aagctcgaagcgaagctggaagacacactggaagcctggaagcgaagcctcgaactcc 3446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 591 CGATTAGAAATTAAGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 650

OY 3447 ctctgagcagctgctcaccgctgctcgcgagctcgaagcagcattgaagaacaagaacta 3506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 651 ATCAGAGAGCTTACACCGATTTTGAAGAATTTAAAGAACTTAATAAAGATTAAGACTA 710

OY 3507 caccgcaata 3517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 711 TACCGGAAAAA 721

```

RESULT 28

```

B13042/c
LOCUS      B13042      1147 bp      DNA      linear      GSS 14-MAY-1997
DEFINITION T30M24-Sp6.1 TAMU Arabidopsis thaliana genomic clone T30M24, DNA
SEQUENCE.
ACCESSION  B13042
VERSION    B13042
KEYWORDS   B13042.1 GI:2094174
SOURCE     GSS.
ORGANISM   thale cress.
            Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 1147)
AUTHORS   Feng,J., Dwyer,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
TITLE      BAC End Sequences at ATGC
JOURNAL    Unpublished (1997)
COMMENT    Other GSSs: T30M24-Sp6, T30M24-Sp6.2
            Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: Sp6
            Class: BAC ends
            High quality sequence start: 289
            High quality sequence stop: 292.
FEATURES   1..1147
            location/Qualifiers
            source
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /clone="T30M24"
            /clone_11b="TAMU"
            /sex="hermaphrodite"
            /note="Vector: BelobACII; Site_1: HindIII; Site_2: HindIII
            ; Produced by Rod Wing"
BASE COUNT      3 a      4 c      46 g      553 t      541 others
ORIGIN
Query Match      1.6%; Score 78.8; DB 12; Length 1147;
Best Local Similarity 23.8%; Pred. No. 3.3e-05;
Matches 251; Conservative 0; Mismatches 805; Indels 0; Gaps 0;

```

```

OY 695 acacatcagagacaatgtgggaagatcggaagtattatataaagaataagaagacca 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1145 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1086

OY 755 tcgagaacattacgagctgacgaagatccaaagaacatagacaagaataagaatg 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1085 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1026

OY 815 caaccagagagagaagaagaagtgctgaccagcccaagcagactgctccactata 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1025 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 966

OY 875 acaacagcttgaagaagcccaataacctcatcagctgtagagaagcgcatagacacc 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 965 NANANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNA 906

OY 935 tcaagaagaatgaataatcaagaactgctcgacaagatgaatgaatgaatcctc 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 905 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 846

OY 995 cgccagcgaactctggagacacccttaacacgctctgtagacaagaagaagaatagag 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 845 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 786

OY 1055 agcagagaagaagatcaagaagatcgccaacacattgaagttcaacatagattctct 1114

```


COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0080 row: C column: 24
Seq Primer: CGTGTAAACGACGCCACT
Class: plasmid ends
High quality sequence stop: 540.
Location/Qualifiers

FEATURES

SOURCE

1. 540
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C2M080C24"
/clone_lib="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42hv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 284 a 34 c 131 g 51 t
ORIGIN

Query Match 1.5%: Score 73.8; DB 12; Length 540;
Best Local Similarity 49.9%: Pred. No. 0.00028;
Matches 186; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

OY 731 atataaaagaataagacacatcgagaacatgaagctgacgaagaatccaaa 790
DB 115 AGATGAAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGA 174
OY 791 agaccatagacaataataagatcaaccaggaggaagaagaagtgtaccagg 850
DB 175 AGAAGAAGAGAAGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 234
OY 851 cccagtcgacctgtccatctatacaaacagcttgaagaagccatacctcatcagc 910
DB 235 AGAAGAAGAGAAGAGAAGAGAAGAGAAGAGACTGAGAAGAGAAGAGAAGAGA 294
OY 911 tactgagaagcgcatagacacccctcaagaagaatgaatatcaagaactgtcgaca 970
DB 295 AGAAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGACTGAGAAGAGAAGAGA 354
OY 971 agatlaatgaatlaagaatccctccgacgaactctgggaacacccctaacacgtgc 1030
DB 355 AGAAGACTGAGAGAAGAGAAGAGACTGAGAAGAGAAGAGAAGAGAAGAGAAGAGA 414
OY 1031 tggacaagaacaagaagatagaggagcaggaagaagaatcaagaagatcgccaaaccca 1090
DB 415 AGAACAAGACAAGAGAAGAGAAGAGAAGAGAAGACAACACCAATCAAAAAGA 474

OY 1091 ttaagtcaacat 1103
DB 475 GCCAGACATAT 487

Search completed: September 20, 2002, 04:11:48
Job time: 14247 sec

THIS PAGE BLANK (USPTO)

Query Match 100.0%; Score 4940; DB 6; Length 4940;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4940; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagcgatgaataatcattctctctctgtctcatcttcttcttcatcact 60
Db 1 CGCACCGGATGAAAAATCATTTTCTCTCTGTTCAATTTCTGTTTATCATCAATACT 60

Qy 61 cagtgagtaaccagaatctctatcagagctgtgttaagaacttggaagat 120
Db 61 CAGTGGGTGACCCACGACATCTTACAGACTGTGTTAGAACTGGAAAGTTTGGAGAT 120

Qy 121 gccgtcttaaccgatacagccctgtctcagaagaagaatgggtctgaatgaaggagc 180
Db 121 GCCGTCTTACCGGATACAGACTGTTCCAGAAAGAGAGATGGTGGCAATGAAGGAGC 180

Qy 181 agtgcacagcgcttcaaacagcaaccggttctaaaggtctgtgctagcgtgtgc 240
Db 181 AGTGCACAGCGCCCTTACAAACACACACCGGTTCTAAAGGTTCTGTGTACCGGTGCG 240

Qy 241 tccggtgggtctgtgtgctcttgagggttccgtctccgagcgagcgtggtcagat 300
Db 241 TCCGGTGGGTCTGTGGCTCTGGGGGTTCCGTGCTCCGGCGACCGTGGCATACGT 300

Qy 301 ggcctcagtgacagcgcggttccggaagacagtcgaagaaccaatccatctgaacaactc 360
Db 301 GCGCTAGTGGCAAGCGGGGTTCCGGGAACAGTGAAGAACCAATCATCTGCAACACTCT 360

Qy 361 agagatctcagagcgaagctctcagcgaacttaagcagccagatggagaactctctc 420
Db 361 AGGATTCGACGCCCAAGTCTACGCCGACTTAAGCACCGAATGAGAAATCTCTCTCT 420

Qy 421 accatcaagagcttgaagtaaccacagctgtgtcagctcactaatacattgctgacactg 480
Db 421 ACATCAAGAGAGCTGAAGTACCACAGTTGTGTGACCTCATATCATATGACGACCTG 480

Qy 481 tgtgataacaatcattcgctcaaatatctgaattgaaggttgaagaagatcaatgaactc 540
Db 481 TGTGATAACATTCATGGCTTCAAAATATCTGATTGACGGTTACAGAGATCAATGAACTC 540

Qy 541 ctgtacaagttgaattctcactcgaactgtcgaagcgaactgaatgaactgtcttcgac 600
Db 541 CTGTACAAGTTGAATTTCTACTTCTGACTTGTCTAAGGCCCAACTGAAATGCTTGGGCC 600

Qy 601 aatgactatgtcaaatctcattcaatttgaagatcagagccaagcaggttgagcgtatg 660
Db 601 AATGACTATTGTCAAAATTCATTCATTTGAAGATCAGAGCCAAACGATTTGGACGTATTG 660

Qy 661 aagaagtgtgtctcgtgatatcgcaagcctctcgacaacaatcaagaagcaatgtgggaag 720
Db 661 AAGAAATTTGTTCTTCGATATTCGAAGCCCTCTGACAACTCAAGGCAATTTGGGGAAG 720

Qy 721 atggaagattatataaataaagaataagaagacatcgagagacattacgagctgactgaa 780
Db 721 ATGGAAGATTATTATTAAGAAATTAAGAAACCATATGAGAACATTTAAGCGCTGATCGAA 780

Qy 781 gaatccaaaagaagccatagacaataaataagaatgcaaaccaaggaaggaagaagaag 840
Db 781 GAATCCAAAAGAGCCATAGCAAAAATTAAGAATGCAACCAAGGAGGAGAAAAGAGAG 840

Qy 841 ttgtaccagggccagtaagcctgtctcatctataaacaagaagcttgaagaagccataac 900
Db 841 TTGTACCAGGCCAGTACGACCTGTCCATCTATTAACAAACAGCTTGAAGAGCCATTAAC 900

Qy 901 ctcatcagcgttacttggaagcagcatagacacccctcaagaagatgaataatcaagaag 960
Db 901 CTCTATACGCTTACTGAGAAAGCCATATGACACCTCTCAAGAAAGATGAATAATATCAAGAA 960

Qy 961 ctgctgcacaagaataatgaatlaagaatcctcgcgccaagcgaactctggaaacacct 1020
Db 961 CTGCTGCACAAGATTATGAATTAAGAATCTTCGGCGACCAACTCTGGGAACACCCCT 1020

Qy 1021 aacacgtctgtgcacaagaacaagaatagaggacgagagaagaatcaagaagatc 1080
Db 1021 AACACCTGCTGACAAAGAAACAGAGATAGAGACACAGAAAGATCAAAAGAGATTC 1080

Qy 1081 gccaaaaccatttaagttcaacatagattctcttcttaacgattcccttgagctggagatc 1140
Db 1081 GCCAAAACATTAAAGTTCAACATAGATCTCTCTTACTGATCCCTTGAAGCTGGAGTAC 1140

Qy 1141 tacttgagagagaagaataagaatatagacatctccgcaagctgagaacaagagatca 1200
Db 1141 TACTTGAGAGAGAAGATTAAGATATAGACATCTCCGCCAAATCGACAAAGAAATCA 1200

Qy 1201 accgaacctaatgaatatacccaatgggtgtgcagtlaccctctgtctatacgaatcaac 1260
Db 1201 ACCGAACCTAATGAATATTCCAATGCTGTGACCTACCTCTGCTATTAACGATATCAAC 1260

Qy 1261 aacgctctcaagagctcaatagcttcggtgtgacttgatttaaccccttgattataagaa 1320
Db 1261 AACGCTCTCAACGAGCTCAATACCTTTCGGACTTATTAACCCCTTTCGATTAATACGAAA 1320

Qy 1321 gaacctctaaagatactcaacagacaatlgagaagaagaatttcaacgaatcaag 1380
Db 1321 GAACCTCTAAGATATCTACACAGACATGAGAGAAAGATTATCAACGAAATCAAG 1380

Qy 1381 gagaagaatcaaaatctgagaagaagaatlgagatgacaagaagaattcaagaagcgc 1440
Db 1381 GAGAAATCAAAATTTGAGAGAAAGAAATTTGAGAGTGAACAAAGTTTACGAAGACCCG 1440

Qy 1441 agcaaaagcttaagatctcaactaaagatgaagaagctgtctgaagagatctatgat 1500
Db 1441 AGCAAAAGCTTAACATATCTACTTAAGAGTATGAAGCTGTGACAGATCTATGAT 1500

Qy 1501 tccaaatcaacaataacatcgacctgacaaactctgaagaataatgaatgggaaacggttac 1560
Db 1501 TCCAAATTCACAAATTAACATTCGACCTGACCAACTTGAGAAATGATGGGAAACCGTAC 1560

Qy 1561 tcttaacaagtgagaagaactgacacacccaataaacttggatccctatgagaatttcaag 1620
Db 1561 TCTTACAAGTGAAGAACTGACACACCAATTAATCTTTGCACTTATGAGAAATTTCTAAG 1620

Qy 1621 cataacttgagaagctccaccaagctcttaagtataatgagaagctatctctgcggaac 1680
Db 1621 CATATCTTGAAGAGCTCACCAAGCTCTTAAGTATATGAGAGACATATCTCGCGGAAC 1680

Qy 1681 atgttctggaagaagaactaaagtattcaagaatctcataagtaagaatcgaaacgcag 1740
Db 1681 ATGTGTGGAGAAAGAACTAAAGTATTACAAAGATCTCATAGTAAGTACGAAACGAG 1740

Qy 1741 atcgagaagcgtgtgtgagaacattgaagaagatgaagaacagttgtttgagaagaagatt 1800
Db 1741 ATCGAGAAGCCTTGTGAGAACATTAAAGAAAGATGAAGAACAGTTGTTGAGAAAGAGATT 1800

Qy 1801 acaaaagacgaataataaacagaatgaagaagatcctgagagctcagatatgttlaagtc 1860
Db 1801 ACAAAAGACGAATAATAAACAGATGAAGATCTGTGAGGTCTCGCATATTGTTAAAGTC 1860

Qy 1861 caagtgcagaagtgctcctcatgacaagaattgattgaaactcaagaagactcaactcatt 1920
Db 1861 CAAGTGCAGAGAGTGTCTCTCATGAAACAAGATTGAGAACTCAAGAAAGACTCAACTCAATT 1920

Qy 1921 ctgaaagaagtggaagttaaacaataatagatgtccgaatgtatataagcagagatt 1980
Db 1921 CTGAAGAAGCTGGAGTTTAAACATTAATATACATGTCCGGAATGTATTAAGCGAGGAAT 1980

Qy 1981 aagcaggaacataactacactcgtactcaagaagaagatagacaaactgaagaatgtctc 2040
Db 1981 AAGCAGGAACCATACTACCTCATCGTACTCAAGAAAGATGAGACAACTGAAGAGCTTC 2040

Qy 2041 atgccaagaatlgagaagcctgtatcaacgaagaagaagaataatgaactgaagaagcag 2100
Db 2041 ATGCCAAAGTGCAGAGCCTGATCAACGAAGAAGAAACATATTAAACTGAAAGACAG 2100

Qy 2101 tcaagataaactcgaagccttccacagaagagagataacggagaagcgtaccacccaagccc 2160

|||||
Db 2101 TCAGATACCTCGAGCCTTCCACAGAGAGATACCGGACGAGCTACCAACCAAGCC 2160
Qy 2161 ggaacaagccgggttcagctctcgaaggcgatagctgcaagctcaagcaagaagcag 2220
Db 2161 ggacaaacggccgggttcagctctcgaaggcgatagctgcaagctcaagcaagaagcag 2220
Qy 2221 aagaagagacagcctccagctccagctccagctccagagagctcaagctccagctccag 2280
Db 2221 AAGCAGCAGACCTCCAGGCTCCAGGCTCCAGGCTCCAGGCTCCAGGCTCCAGGCTCCAG 2280
Qy 2281 ccaccagctccctgtaataaagaacccgagaatgtaagcaagctggaagctggaag 2340
Db 2281 CCACAGCTCTGTGAATTAACAGACCGAATGTACAGCAAACTGACTGACTGACTGAGAG 2340
Qy 2341 ctctatgagctccctggaataactctcaactctgcgaataatactctgctctcaagc 2400
Db 2341 CTCTATGAGTTCCTGAATTAACATCTCTACATCTGCAAAATATCTCTGCTCTCAGAGC 2400
Qy 2401 actatgaacggaagatctcttaaacagtaacagtaacgaaggaaggaagtaaacgt 2460
Db 2401 ACTATGAACGGAAGATTTCTTAACAGTACAGATTAACCAAGGAAGAGAGATTAAGCTG 2460
Qy 2461 tccctctgtaaccagctggaacctgctgtcaataaccagaaacaactcccgtaagat 2520
Db 2461 TCCTCTGTGATCCACTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520
Qy 2521 tctatgttcgtaagctcaacaactctctctctcaactgctcaagagatataatgaag 2580
Db 2521 TCTATGTTCGATAGCTCTGCAAAATTTCTCTCTCAACTGTCTGGAATATATATGAGAG 2580
Qy 2581 gagatgtctgcaacctgtataaactcaagaacgaacgaagatgaagaaacctctgag 2640
Db 2581 GAGATGTCTGCAAACTTTATTAACCTCAAGACAAAGCAAGATTAAGAACTTTGAGAG 2640
Qy 2641 gaaagtaagaagctcccaacctctgttaaaactctctctccagctccagctcaaacctg 2700
Db 2641 GAAGCTAAGAGGCTCCAGCTCTGTAAACTCTCTCTCAGCTCCATGCAACCACTG 2700
Qy 2701 tctctcaacctcaagaagaacccggaagtgaagcgtcaagcagaacacctctcaactgc 2760
Db 2701 TCTCTCAACCTCAAGAACCCGAAATGAGCGCTTAACGACGACCTCTCACTCGAGC 2760
Qy 2761 aaccttaaacactcaactgaacctgttgaagacacctgctctcgtggaagaaatgaag 2820
Db 2761 AACCTTAATTAACCTCACTGAATCTGTTGAGAACATCTCTCTCTGCGAAGATTAAGAC 2820
Qy 2821 actcaacaagaacttatgtgaacagaacatcgctcgaagaactctcaagagaatatactgaa 2880
Db 2821 ATCTAACCAAGAACTTATGAGACAAATCGTCCGAACCTCTACGAGAACATACTGAAA 2880
Qy 2881 gaacagcagaacatctctataagaagagcttcaactaactcgtggaatctcaagccgaat 2940
Db 2881 GACAGCGACAACTTCTAATTAACGAGAGCTTCACTAACTCTGGAATCTTAAGCCCATAT 2940
Qy 2941 atcaactctctcaagatgaatctcaacgtaagaagcgtggaagagagataataagctg 3000
Db 2941 ATCAACTCTCTTAACGATGATCTTAACGTAAGAACTGGAAGAGACATTAATTAAGCTG 3000
Qy 3001 aagaagaacactgcaactgagctcgaacctgtacaacaagtaacaactgaaactggaaga 3060
Db 3001 AAGAAGACACTGCAACTGAGCTGAGCTGTACAAACAGTAACAACTGAACCTGAGAGA 3060
Qy 3061 ctctctgaagaagaagaagcagctcggaagataagaatgaagaatgaagctg 3120
Db 3061 CTCTCTGCAAGAAGAAACAGTGGGCAAGATTAAGATGCAAGATTAAGATTAAGCTG 3120
Qy 3121 ctcaagagcagcttgaagaagaacactcaactcaactgaacaaactcgaagcagtaactgag 3180
Db 3121 CTCAAGAGCAGCTTGAAGAAAGCAACTCACTGAAACATCTCCGAACACGTAAGCTGAG 3180
Qy 3181 aaactctcagtgctctcaacaagaagaagccgagatcgccgagacagagaacact 3240
|||||

|||||
Db 3181 AACTTCTCAGTGTCTTCTCAACAGAAAGAGAGCCGAGATCGCCGAGACAGAACT 3240
Qy 3241 ctggaagaacccaagaatctctctcaaacactacaagaagctctgtaagatatataatgagc 3300
Db 3241 CTGGAAGAACCCAAATTTCTCTCAAACTCAAGAAAGCCCTGCTCAACTATTATTAATGAGC 3300
Qy 3301 gactctctccctgaagaacctctccgagagagaacatcaagaacccgagataactagcgc 3360
Db 3301 GAGTCTTCTCTCTCAACACTCTCTCCAGAGAGACATCCAGACCGAGATTAATTAAGCC 3360
Qy 3361 agctcgaacactcaagctcgtctcaagctcgaaggcaagctgaaagcaaacctgagc 3420
Db 3361 AGCCTCGAAGAACTTCAAGCTCTGCTTAAGCTCGAAGCAAGCTGAAGCAACCTGAGAC 3420
Qy 3421 ctggaagaagaagcagtaagctacactctctgagcagctgcaactctgagcagctg 3480
Db 3421 CTGGAAGAAAGAACTCAGCTACCTCTGAGCGGAGCTGCAATCATCTATGCGGAGCTG 3480
Qy 3481 aagaagctcatlaagaagaagaactcacacggcaatagcccaagcgaataataacagac 3540
Db 3481 AAGAAGCTATTAAAGAAACAGAACTACACCGCAATTAACCCAGAGGAGATTAATTAAGAC 3540
Qy 3541 gtgaataacgcaactggaatcttaagaagctcctgctgctgaaggaaacagatgctgcact 3600
Db 3541 GTGAATTAACGCACTGGAATCTTACAAAGATTCCTGCTGGAAGAAACAGATGCTGCCACT 3600
Qy 3601 gtggtgtctgtaatctgctctccgacacactgtagagagcttcaacctgaagccctgcact 3660
Db 3601 GTGGTGTGAATCTGCTCTGCGACACTGGAAGCTGTCAACCTTAAGAACTCTGCACT 3660
Qy 3661 actcatgctcgagccgaatcccaatacaatcaacacatccgaagctcgaagatgagctc 3720
Db 3661 ACTCATGCTCGAGCGGAGCTCCAAATTAATTAACATCTCAAGAGCTGAGATTAAGCTC 3720
Qy 3721 gatgacgtcatcatgtgctcctacactcctcgagagcgaaggagactacgataacctgcgc 3780
Db 3721 GATGACGTCAATCATGTGCTCTATCTTCCGAGAGCCGAGAGGAGCTAGAGACTCTCGGC 3780
Qy 3781 caggtgtcaccagctggaagcgtgaactctctcgtgagatgataactcttccaaaact 3840
Db 3781 CAGGTGTCCACCGGTGAGGCTGTCACCTCTTCCGATTTGAATCTGTCCAAATTC 3840
Qy 3841 gagaacgaatacgaagctgctctatctgaaacctctggaagcgtctataagctctcctcaag 3900
Db 3841 GAGAAGCAATTAAGAACTCTCTATCTGAACCTCTGCGAGCGCTTATTAAGCTCTCTCAAG 3900
Qy 3901 aaacagctggaagaataacgctgaactcctcaatgtaacgctggaagacatctggaagc 3960
Db 3901 AAACAGCTGGAATTAAGCTGATGACCTTCAATGTCACGTAAGAGACATCTGAAACAGC 3960
Qy 3961 cgtcttaataagaagaataatctcaagaagctcttggagagcagctgataatccctataaa 4020
Db 3961 CGCTTTAATAAGAGAGAAATTTCAAGAACTCTTGGAGACGACTTGAATTCCTATTA 4020
Qy 4021 gacctgacctcctcaactagctgttgaaggaccatacaagcttccctcaataagaagaag 4080
Db 4021 GACCTGACCTCTCTCAACTAGCTGTGTCAAGAACCATTAAGCTCTCAATTAAGAGAGAG 4080
Qy 4081 aggataaactctgctctagttacaactatacaagaactccatcgacacccgataatcaat 4140
Db 4081 AGGATAAATTTCTGTAGTTAACAATTAATTAAGAGACTCCATCGACACCGATTAATCAAT 4140
Qy 4141 ttgcgtatgtagtgcctgggttatcaagaatcctgagcgaagaaataagaatctgagctc 4200
Db 4141 TTGCGTAAATGATGTCTGGGTATTTAACAAGATCTCGACGGAATAATTAATCTGACCTT 4200
Qy 4201 gactctatataaagatatatacaacgataagaagcgaagctgaaataatctgacctc 4260
Db 4201 GACTCTATTAAAGTATATCAACGATTAAGCAAGCGGAGATTAAGAAATATCTGCCCTTC 4260
Qy 4261 ctgaataacatcgaaacccgttacaagaagctgaagcgaacaaactcgacctctcgtaact 4320
Db 4261 CTGAATAACATCGAAACCTGTACAGACAGTGAACGCAAAATTCGACCTTCTGTATTT 4320
|||||

Oy	4321	caccctggaggccaagtctcctaactatacttaacggaagagcaatgttgaattaaatc	4380
Db	4321	CACCTGGAGGCCCAAGTCCCTCAACTATATCGAAGAAGGCAAAGTGGAAATTAAAAATC	4380
Oy	4381	aaggagctgaacctactcaaaccaatlccaaagaacaagcttgycagattccaagaaaataac	4440
Db	4381	AAGGAGCTGAACCTACTCTAAAACAATTCACAGAACAGTGCGGAGATTTCAGAAAAATTAAC	4440
Oy	4441	aatttcgtcggaattgcagaactgtctctacgaattataaccacaacaatctcctgaaccaag	4500
Db	4441	AATTTCTGTGGAATTCAGAGACCCTGTCTACCGATTTATAACCAACAANAATCTCTGTGACCAAG	4500
Oy	4501	tctctgcacactcgggatgtgtctgcgaaaaaccgcgcgaacaaacagttgctagcaaatctgc	4560
Db	4501	TTTTCTGTCCACTCGCATGTGTTCAAAAACCTCGCCAAAACAGTCTGAGCAATTTGCTTC	4560
Oy	4561	gacggaacaacttcgaaggcatgtctgaacatctccacagcaccaatgcgtgtgaagaacagttgc	4620
Db	4561	GACGGCAACCTGCAGGGCATGCTGAACAATCTCCACAGCACCAATGCGTGAAGAAACAGTGC	4620
Oy	4621	ccccgaatatagcgcgtgttctcaggcaatcttgacgcagcgcgcgaagaattgcgaatgtctctgc	4680
Db	4621	CCCCGAATATAGGGCGTGTTCAGGCATCTGACGAGCGAGCGGAAGTAGTGCATGTCTCCG	4680
Oy	4681	aactcaacaagaagagatgaatgcgtgtgagaaccacaacccctacactgcgaatgaaac	4740
Db	4681	AACCTCAAAACAAGAAAGAGATAGTGCCTGAGGAACCCAAACCTTACCTGCAGATGAANAAC	4740
Oy	4741	aatggcgggtgtgacgcgcgaatgcataatgcacgcgaagaaagacagcggctctacaaggaaag	4800
Db	4741	AATGGCGGGTGTGAGCCGCATCTAAATGCAACCGAGGAAGACAGCGGCTCTTAACGGAAAG	4800
Oy	4801	aaatcacatgtagtgtactaagcccagctcctatccactcttcgaacgggaattttttgc	4860
Db	4801	AAAATCACATGCGAATGTACTAAGCCCGCACTCTTATCCACTTTTGACGGGATTTTTTGC	4860
Oy	4861	tccagctcctaattcctctggcatactcctctcgtcgtacatccatcgtatcgtatcgtacagc	4920
Db	4861	TCCACCTCTAAATTTCTGTGGCATCTCTCTTCCTGATCCATCATGCTGATCCTGAACAGC	4920
Oy	4921	ttcatcataagatcgatgg 4940	
Db	4921	TTCAATCTAATATGATCATGG 4940	
RESULT	2		
LOCUS	CVEJ13294	4940 bp	DNA linear SYN 24-FEB-1995
DEFINITION	Cloning vector pBSK+-MSP-Is/FCB.		
ACCESSION	AJ131294		
VERSION	AJ131294.1 GI:4455908		
KEYWORDS	cloning vector; major surface protein-1; msp-1 gene.		
SOURCE	Cloning vector pBSK+-MSP-Is/FCB.		
ORGANISM	artificial sequence; vectors.		
REFERENCE	1 (bases 1 to 4940)		
AUTHORS	Bujard,H.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-DEC-1998) Bujard H., ZMBH - Zentrum fuer Molekulare Biologie, Universitaet Heidelberg, Im Neuenheimer Feld 282, 69120 heidelberg, BW, GERMANY		
REFERENCE	2 (bases 1 to 4940)		
AUTHORS	Pan,W., Ravot,E., Tolle,R., Frank,R., Mosbach,R., Turbachova,I. and Bujard,H.		
TITLE	Vaccine candidate MSP-1 from Plasmodium falciparum: a redesigned 4917 bp polynucleotide enables synthesis and isolation of full-length protein from Escherichia coli and mammalian cells		
JOURNAL	Nucleic Acids Res. 27 (4), 1094-1103 (1999)		
MEDLINE	99128299		
FEATURES	Location/Qualifiers		
SOURCE	1..4940		
	/organism="Cloning vector pBSK+-MSP-Is/FCB"		

source	sig_peptide	gene	CDS
/db_xref="taxon:85834"			
/focus			
1..4940			
/organism="Plasmodium falciparum"			
/strain="FCB-1"			
/db_xref="taxon:5833"			
10..4929			
/gene="msp-1"			
10..69			
/gene="msp-1"			
10..4929			
/gene="msp-1"			
/function="surface protein"			
/codon_start=1			
/evidence=experimental			
/product="major surface protein-1"			
/protein_id="CAB36903.1"			
/db_xref="GI:4455909"			
/translation="MKIIIFLCFLPIINOCYTHESYOELVKKLEALDEAVLTGVS			
LFQEKWMLNGTSGTAVTTSTPSGKSVASGGGGSVASGGSVASGGSVAS			
GGSSNRTSPDSNDSDDAKSVADLHRYNVLITIKELYPOLFDLTHMLTLC			
IHGRKYLLDGYEELINELLYKLNFLNYEDLRALNDVCAADYQIPNLKIRANL			
KLVGKRPDLNIDKNDVKMEDYIKKNKTTENINELIEESKTIIDKNKATPEEK			
KLYOAYDLSTYKNOLEAHNLISVLEKIDITLKNELIEESKTIIDKNKATPEEK			
NTPTLIDKNKITEHEKEIKETAKTIKFNIDLSFTPLDELEYLREKNKIDIAK			
ETKSTPEENKPYTPPLSLNDINNAINELNSGDLINPDYIKESKNITPNEAK			
KFINDEKIKIEKKIESDKSTEDRSKSLNDITKEKRLNLTDSKFNNDITLN			
FEKMKGRYSYKVERKLTHTNFASYENSKNLKLTALKYMEDYSLKNIVOKELK			
YKNLSIKIENIEETLVENIKDEQLEKKITJEDENKDEKILEVSDIVKVOYVLL			
MNKIDELKTIQTLINVELKNHINIVPNYKQENKQEPYLLIVLEKIDLEKMPK			
SLINDEKKNITTEGSDNSESTGEITGATTPRGQOAGSALGSDSOAOAOKOKO			
QPEVLPVPEAKAOVPPRPAPVNNKTEVSLDYLEKLYEELNTSYICHKILVSH			
MEKTLKQYTKTDEESKLSQCDPLDLFLPNQNNIPWMSFSDLSNLSQLPMEIT			
KEMCNLYKTKDNDRKMLLEAKRVSIVATLSSSQKQPLSTPQDPEVASNDIT			
HSTINDNSLKLEENILSLGKNKITVQELIGQSENEFEKLIKDSDFYENSFN			
SKADINDSLINDESRKLEEDINKLTKTQFLDYNKYKLEKRLPEKKTVKYM			
QIKKTLTKLEQESKLSLNNPKHYLQNFVYFPNNKKEALIAETENLLENKILKH			
KGLVKNYENSPKLTLEESIIOTEDNVALSENFVKSLGKRLKDNLENKKLSY			
SSGLHLIAELKPVTKNNYNGNSPENNPNVNALESYKLPBGDVAIVLSESS			
DALVPSQPKRPASTHVGAESNTITTSQVDEVDVIVIPIFGSEEDYDLGOVYTG			
EATVPSVIDNLISKILENEYEVLYLKLQAGVRSLLKOLENNWTFNNVNDILS			
KRENFKNVLESDILPYKDLTSSNVVKKPYFLKNEKKDKFLSSNYIKSIDPND			
ANDVLYGKILSEKSDLSIDIKKYINDKOENKRYEPLNINETLKYTVNDRID			
ITHLEKVLNTYKESNVEVKIKELNTLYTIDODKLADFKVKNKQCPANGSCF			
LTKPLSTGMVFNENAKTYLSNMLDGNLOGMATNISOHCYKQCPANGSCFPLD			
CKKLNLTQKQEDKVENPNPCENENNNGCDADACTEEDSSNKKITTECTKPDST			
LFDPISFSSNFIPLIIMLILYSPI"			
BASE COUNT 1669 a 1156 c 1060 g 1055 t			
ORIGIN			
Query Match 100.0%; Score 4938.4; DB 12; Length 4940;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
1 ccgaagcgatataaatatatttcttcctctgttcattctgtttttataatcaatc 60			
111 60			
1 ccacagcgatataaataatctttttcttccctctgttcattctgtttttataatc 60			
111 60			
61 cagtcgctgaaccacgaatccctatacgaagcctgtgttaagaacctggaacct 120			
111 120			
61 cagtcgctgaaccacgaatccctatacgaagcctgtgttaagaacctggaacct 120			
111 120			
121 gccgcctctaccgaatacagcctgttccagaagagaaatgtgtcctaataaagg 180			
111 180			
121 gccgcctctaccgaatacagcctgttccagaagagaaatgtgtcctaataaagg 180			
111 180			
181 agtggcagcgccgttacaacacgaaccgggttctaagggtctgtgtgctaggc 240			
111 240			
181 agtggcagcgccgttacaacacgaaccgggttctaagggtctgtgtgctaggc 240			
111 240			
241 tccggtgggtctgtgtgctctgtggggttccgtcgccctccggcggaagcttgatc 300			
111 300			

Dh 241 TCCGGTGGGCTGTGGCCTTGGGGGTTCCGTGCCCTCCGGCGGAGCGTGGCATCAGT 300
Qy 301 ggtcaatggaagaagcgagcttcggygaacagtcgaagaacaatcactctgaacactc 360
Dh 301 GGTCTAGTGGCAAGGGGGGTTCCGGGAACAGTGCAGAACCAATCCATCTGACAACTCT 360
Qy 361 agcgaattccgaagcccaagtcctacgcgcgaacctcaagcaccgaatgagaacaatctctc 420
Dh 361 AGCGATTCGAGCGCCAAATCTCTACGCCGACCTCAAGCACCGAGTGAAGAACTATCTCCTC 420
Qy 421 actatcaagaagcggaagtaaccacagttgttcgaacctcactaatacatgtctgacacty 480
Dh 421 ACTATCAAGGAGCTGAAGTACCCACAGTGTTCGACCTCCTCAATCAATATGCTGACACTG 480
Qy 481 tgtataacatcatcagcttcaaatatctgaattgaacggtttacgaagaatcaatgaactc 540
Dh 481 TGTGATACATTCATGGCTTCAAAATATCTGATTGACGGTTACGAAGAAATCAATACATC 540
Qy 541 ctgtacaagttgaatttctacttcgaacttgcctgaagggccaactgaatgaacgtttgcgc 600
Dh 541 CTGTACAGGTTGAATTTCTACTTCGACTTGCCTAAGGGCCAATGAATGACGTTTGGCGC 600
Qy 601 aatgaactatgccaatccatccaatttgaagatcagaaccagatlggaacatltg 660
Dh 601 AATGACTATTTGTCAAAATTCATTTCAATTTGAAGATCAAGGCCAACGAGTTGGACGATTTG 660
Qy 661 aagaagtgtgtcttcgagatcgcgaagctctgcgaacaatccaagaagaatgtggaag 720
Dh 661 AAGAAGTGTGCTTCGGATTCGCAAGGCTCTCGACAACATCAAGCAAAATGTGGGAAG 720
Qy 721 atggaagaatlatataaagaataaagaaccatcgaagaacatlaagagctgtacgaa 780
Dh 721 ATGGAAGATTATTTAAAAAGATTAAGAACACATCGAAGACATTAAGAGCTGATCGAA 780
Qy 781 gaatccaaaagaaccatagacaaaaaataagaatgcacccaaggaagaagaagaag 840
Dh 781 GAATTCAAAAAGACCATGTGACAAAAAATTAAGATGCAACCAAGAGAGAAAGAAAGAGAG 840
Qy 841 ttgtaccagggccagtaagacactgtccatctatacaacagcttgaagaagcccatlac 900
Dh 841 TTGTRCCAGGGCCCACTACGACCTGTCCATCTTAACCAACAGCTTGAAGAAAGCCCATAC 900
Qy 901 ctcatcagcgtacttggagaagcgcatagacacccctcaagaagaatlgaaatatcaagaa 960
Dh 901 CTCATCAGCGTACTGAGAGAGCGCATAGACACCCCTCAAGAAAGAAATATATCAAGAA 960
Qy 961 ctgctgcgaagaatataagaatctcccgcaagcaactctvggaaacacccct 1020
Dh 961 CTGCTCGCAAGATTAAATGAATTAAGAAATCTCCGCCAGCCAACTGTGGGAAACCCCT 1020
Qy 1021 aacacgctgtcgtgacaagaagaagaagaatagaagagcagaagaagaatcaagaagatc 1080
Dh 1021 AACACGCTGCTGACAAACAGAGATAGAGAGCAGCAAGAAAGACATCAAAAGATATC 1080
Qy 1081 gccaaaacatlaagttcaacatagatctctcttacttgaatccctctgaagctgtgaagtc 1140
Dh 1081 GCCAAAACCATTAAGTTAACAATAGATTCTCTCTTACTGATCCCTTGAGGTGAGATAC 1140
Qy 1141 tacttgagaagaagaataagaatatagacatctccgcgaagtcagacaagaagatca 1200
Dh 1141 TACTTGAAGAGAAATTAAGAAATATAGACATCTCCGCAAAAGTTCGAGACAAAGAAATCA 1200
Qy 1201 accgaacctatgaataatcccaatgtgtgaagctacacctgtgtctataagaatatcaac 1260
Dh 1201 ACCGAACCTTAATGAATATCCCAATGAGTGTGACGATACCTCTGTCTTATTAACATATCAAC 1260
Qy 1261 aacgctctcaacagagctcaatagctctcgttgaacttgaatlaaaccctctgaatataga 1320
Dh 1261 AACGCTTCAACGAGAGCTCAATAGGCTTGGGTGACTTGAATTAACCCCTTGATTAATACGAAA 1320
Qy 1321 gaaccccttaagaatatctacaagacaaatgaagagaagaagtttacaacgaatacaag 1380
Dh 1321 GAACCCCTTAAGAATATCTTACACAGACAAATGAGAGAAAGATTATATCAACCAAAATCAAG 1380

Qy 1381 gagaatccaataatlgagaagaagaanaatlgagagtgaacaagaanaagttacaagaaccgc 1440
Dh 1381 GAGAAGATCAAAATTTGAGAGAGAGAAATTTAGAGATGACAAAGAAATTTACGAACACCGC 1440
Qy 1441 agcaaaagltcaaacgatactcaataaagatatagaagaagctgtgcgaagatctatgat 1500
Dh 1441 AGCAAAAGTCTTAACGATATCTAATAAGATATGAAGAGTGTCTGACAGATCTATGAT 1500
Qy 1501 tccaattccaataataatcagacctgcgaacacttcgagaataatratggaaaaagttac 1560
Dh 1501 TCCAAAATTCACAAATTAACATGACCTGACCAACTTCGAGAAAATGATGGAAAAGGATAC 1560
Qy 1561 tcttaacaagttgagaagaactgcacacccaataataacttgcataccatagaagaattcga 1620
Dh 1561 TCTTAACAAGTGGAGAAATCGACACACATATATACCTTTGCATCTTATGAGAAATTCATAG 1620
Qy 1621 cataactcttgaagaagctcaccacaagctcttaagtalatlgagagactatctctgcgaac 1680
Dh 1621 CATATCTTGAAGAGCTCACCAAGCTCTTAAGTATATGAGAGACTATTTCTCGGGAAC 1680
Qy 1681 attgtgttggagaagaactcaaatgaattatacaagaatctcataagttaagatcgaaaaagag 1740
Dh 1681 ATTGTTGTGGAGAAAGAACTAAAGTATTACAAGAATTCATTAAGTAAATGATCGAAAAAGAG 1740
Qy 1741 atcgagacgctgtgttgaagaacatlaagaagaatgaagaacagttgttgaagaagaagatt 1800
Dh 1741 ATCGAGACGCTTGTGTAAGACATTAAAGAGATGAAGAACATGTTTGTGAGAAAGAAATT 1800
Qy 1801 acaaaaagcgaanaataaaccagaatlgagaagaatcctcgtgagcttcgcgaattgttlaaagtc 1860
Dh 1801 ACAAAGAGCGAAATTAACCAAGATGAGAAATCTCGAGAGCTCCGATATTTGTTAAAGTC 1860
Qy 1861 caagtgcagaaggtgtcctcctaatagaagaatgtgaactccaagaagaactcaactcatt 1920
Dh 1861 CAAGTGCAGAAAGTGTCTCTCATGAACAAGATTGATGAACCTCAAAACATCAACTCATTT 1920
Qy 1921 ctgaagaacgttgaagttlaaacaataatactatcgtgcgcaatagttacaagaagaagat 1980
Dh 1921 CTGAAGAACGTGAGATTAAACATATATACATGTCGCGAATAGTTATTAAGCAGAGAGAT 1980
Qy 1981 aagcagaacatctactaccatcgttactcaagaagaagaatagacaaaactgaagtgctc 2040
Dh 1981 AAGCAGGAACCATACTACTCATCTCATCTCAAGAAAGAGATAGCAAACTGAAATGTGTTT 2040
Qy 2041 atgcccaagttcgaagctctgaatccaagaagaagaagaacatlaaacctgaagaagag 2100
Dh 2041 ATGCCCAAGTCTGAGAGCTCTGATCAACGAGAGAAAGAAACATTAACCTGAAGGACAG 2100
Qy 2101 tcaagtaactccgaagccttccacagaagaagaatgaacccgaagcgtataccacaagccc 2160
Dh 2101 TCAGTTAACTCCGAGCTTCCACAGAGAGAGATTAACCGGACAGGATCCACCAAGCCC 2160
Qy 2161 ggaacacagcgccglttcaagctctcgaagcgatagcgtgcgaagctcaagcacaagaagcag 2220
Dh 2161 GGACAAACAGCGCGGTTTCAGCTCTCGAAGCGGATAGCTGCAAGCTCAAGCAACAGAGCAG 2220
Qy 2221 aagcagcgacagcctccaggtgcaggtgccggttccagaaggttaaaagctcaagtgctaca 2280
Dh 2221 AAGCAGGACAGCCCTCCAGTGCAGTGCCTGTTCCACAGAGGCTTAAGGCTCAAGTGTCTACA 2280
Qy 2281 ccaccagctcctgtgaataaacaagacgagaatgtcgaacaaactggaacttaacttgaagag 2340
Dh 2281 CCACAGCTCTGTGTAATTAACAGACCCAGAAATGTACACAAACTGTGACTTGAAGAG 2340
Qy 2341 ctctaaggttctcgtgaatatacactacatctgcgcaaaatatalactcgtctctacagc 2400
Dh 2341 CTCTATGAGTTCTGTAATACATCTTACATCTGCGCAAAATATATCTCTGCTCTACAGC 2400
Qy 2401 actatgaacggagaagaattcttaaaagttacaagaataaccgaaggagaaggttaaaacty 2460
Dh 2401 ACTATGAACGAGAAAGATTCTTAAACAGTACAAAGATTAACCAAGAGAGAGATTAACCTG 2460

QY 2461 tccctctgataccactgagcctgctgtcaatataccagaacacatactccgtatglat 2520
|||||
Db 2461 TCCCTCTGGATGACACTGGACCGTGTTCAATATTCAGAACAAATCCCGTTATGTAT 2520
QY 2521 tctatgtctgatagctcaacaattctctctcaactgttcaaggatataatgaag 2580
|||||
Db 2521 TCTATGTTGGAATAGCTCTCAACAATCTCTCTCAACTGTCTCAAGATATATGGAAG 2580
QY 2581 gagatgtctgcaacctgtataaactcaagaacagacagaatgaagaacctctggag 2640
|||||
Db 2581 GAGATGCTGCAACCTGTATTAATCAACAGACACAGACAAAGTTTAAAGACCTCTCGAG 2640
QY 2641 gaagtaagaaggtctccactctgtttaaactctctctccagctccatgcaaccatg 2700
|||||
Db 2641 GAAGCTAAAGAGGTCCACCTCTGTTAAACCTCTCTCCAGCTCCATGCAACCACTG 2700
QY 2701 tctctacacctcaagaacagcccggaagtgtgagcgctaaagagacacctcaactgac 2760
|||||
Db 2701 TCTCTACACCTCAAGACAAAGCCGAGGTGAGCGCTAACGACGACACCTCTCACTGACC 2760
QY 2761 aaaccttaactcaactggaactgtttggaacatcctgtctctcggaagaataagaac 2820
|||||
Db 2761 AACCTTAATTAACCTCACTGGAACGTTGAGAACATCTCTGTCTCGGCAAGATTAAGAAC 2820
QY 2821 atctaccaagaacttatggaacgaatcgtccgagaacttctcagaagaatatactga 2880
|||||
Db 2821 ATCTACCAAGAACTTATTTGACAGAAATGCTCCGAGAACTTCTACGAGAAATACGAAA 2880
QY 2881 gacagagacacattctataagagagctcaactcaactcgtgaactctaaagccgagt 2940
|||||
Db 2881 GACAGGACACATTCTATTAACGAGACCTTCACTTCTGTGAATCTTAAGCCGATGAT 2940
QY 2941 atcaactcctttaaagatgaatctaaacglaagaagctggaagagacatacaataag 3000
|||||
Db 2941 ATCAACTCTTAAACGATGAATCTAAACGTAAGAACCTGGAAGAGCAATCAATAAGCTG 3000
QY 3001 aagaagaacactgcaactgagctcgacactgtatacaacagtaacaactgaaactg 3060
|||||
Db 3001 AAGAAACACTGCACTGACGTGACCTGACCTGTACMACAACTGTAACAACTGAAACGGAAGA 3060
QY 3061 cctctcgaagaagaagaagcagtcggaagatgaagatcagatcagaatgaagaagt 3120
|||||
Db 3061 CTCTTTCGAAGAAGAGAGACAGTCGGCAAGTATTAAGATGACATCAAGANCTTGACTCTG 3120
QY 3121 ctcaagagcagcttgaagaacaaactcaactcaactgaaacatccgaacacgtaactg 3180
|||||
Db 3121 CTCGAAGACAGCTGTAAGGCAAACTCACTCACTGCAACATCCGAACACGACTGACAG 3180
QY 3181 aactctcagtgcttctcaacaagaagaagacgagatcgcgagacagagaact 3240
|||||
Db 3181 AACTTCTCAGTGTCTTCAACAAGAAAGAAAGGAGCCGAGATCCGACAGACAGAACACT 3240
QY 3241 ctggaagaacacaaagtcttctcaacaactataaagacctgtcaagtatataatg 3300
|||||
Db 3241 CTGGAGAACACCAAGATTCTTCTCAACACTTACAAAGCCCTGCTCAAGTATTTAATGGC 3300
QY 3301 gagcttctcctctgagaactctctcgaagagagatccagaacggaagtaactag 3360
|||||
Db 3301 GAGTCTTCTCCTGTAAGACTCTCTCCGAGAGAGCATCCAGACCGAGTAACTACGCC 3360
QY 3361 agcctcgaagaactcaaggtcctgtcttaagctcgaagcgaagctgaagaacaacctg 3420
|||||
Db 3361 AACCTTCGAAGAACTTCAAGGTCTCTGTTAAGCTGGAAGGCAAGCTGAAGGCAACCTGAAC 3420
QY 3421 ctggaagaagaagactcagactcactcttaagcgagatcgaactgatacgcgcgag 3480
|||||
Db 3421 CTGGAGAAAGAAAGCTCAGCTTACTCTTAAGCGGACTGATCACTGATCGCGAGCTC 3480
QY 3481 aaggaagtcaataagaacaagaactacacgcgaatagcccaacggaagataataag 3540
|||||
Db 3481 AAGGAAGCTATTTAAGAAACAAGAACTACCGGCAATAGCCCAAGCGAGATTAATACAGAC 3540
QY 3541 gtaataacgcaactggaactctataagaagttctcgtcctgaaggaacagatgtcg 3600
cact

Db 3541 GTGAATAAGCACTGGAATCTTACAGAAGTTCTCTCGTAAGGAACAATGTGCGCACT 3600
QY 3601 gtggtgtcgaatctgtgtccgcgacacacgcggaagcagttcacaactagaagctg 3660
|||||
Db 3601 GTGTTGTGTAATCTGTGGCTCCGACACAGTGAACACTGTCAACTTAAGAAAGCTGTGATCT 3660
QY 3661 actcatgtcgaagccgaagtcacaaatacaatatacacatctcaagaagctgaagatg 3720
|||||
Db 3661 ACTCATGTGGAGCCGAGTCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3720
QY 3721 gatgacgtacatctgtgctctatcttcgagcgaagcgaagaggaactcagatgac 3780
|||||
Db 3721 GATGACGTATCATTTGTGCTATCTTGGCGAGAGCGAGAGAGACATGACATGCTCGGC 3780
QY 3781 caagtgtcaccggtgagctgacactcctcctcggtatgataaactgtgtccaat 3840
|||||
Db 3781 CAGGTGTACCGGATGAGCTGTCACTCTTCCGTATTTATTAATTTGTGCAAAATC 3840
QY 3841 gagaacgaatacgaagtgtctctatactgaacactctggaagcgtctataagttc 3900
|||||
Db 3841 GAGAACGAATACGAAGTGTCTATCTGAACCTCTGCGAGCGCTATATAGTCTCAAG 3900
QY 3901 aaacagctggaataacgttgatgacctccaatgttaacgtgaagaactctgaacac 3960
|||||
Db 3901 AAACACTGGAAGATTAACGTGATGACCTTCAATGTCAACGTGAAGGCAATTTGAAACAC 3960
QY 3961 cgttcaataagaagaagaatctcaagaacgctctggaagcgaactgtatcccta 4020
|||||
Db 3961 CGCTTTAATTAAGAGAAATTTCAAGAACGCTTGGAGAGCGACATGATTCCTTAATA 4020
QY 4021 gacctgaactcctctaaactacgtgttcaagagaccatacaagttctccaataaga 4080
|||||
Db 4021 GACCTGACCTCTCTTAACATCGTTTCAAGAGACCCATTAAGTCTCTCAATAAAGGAAG 4080
QY 4081 agggataaatttctgtctgattacaactatataagagactccaagaccgatatac 4140
|||||
Db 4081 AGGGATTAATTTCTGTGTTGTTCAACTATTAAGAGCTCCATCAACACCGATATCAAT 4140
QY 4141 ttccgtaatgattgtctgtggtatctcaagaatcctgaagcgaataatacaagttc 4200
|||||
Db 4141 TTGCTTAATGATGTGCTGGGGTATTACAAGATCTGACGGAATAAATATGATGACCT 4200
QY 4201 gactctataaagaatatacaacgaataagcgaagcgaatgaataataatctgccc 4260
|||||
Db 4201 GACTCTATTAAAGATATATCAACGATTAAGCAAGGAGATCAAAAATATGCTGCCCT 4260
QY 4261 ctgaataacatgaaacccctgtataagaacagtgaacgaaataatgaacctctg 4320
|||||
Db 4261 CTGAATTAACATGCAAAACCTGTACAAAGACGTAAACGACAAATGCAACCTCTGTAAT 4320
QY 4321 cacttgagcgaagttcctcaactaacttacttaagagaagcgaatgtggaagtta 4380
|||||
Db 4321 CACTGAGAGCCAAAGTCTGCAACCTATACTTACGAGAAGGCAATGTGAAGTTAAAT 4380
QY 4381 aaggaagctgaactcctcaaaacaatccagaagcgtgagattcaagaataatac 4440
|||||
Db 4381 AAGGACTGAACCTCACTCAAAACAATCAAGAACCTGCGCAATTTCAAGAAATAA 4440
QY 4441 aatttcgtcggaattgcaaacctgtctaccgattataacacaacaataatctcg 4500
|||||
Db 4441 AATTTCGTGGAAATTCAGACCTGTCTACGATTATTAACACACAAATCTCTGACCAAG 4500
QY 4501 ttctctgaccatgcaatgttctgaaaaacccctgcgaacaaagctgtgcaactct 4560
|||||
Db 4501 TTTCTCTCCACTGCAATGTGTTGTAAGAAACCTGCAAAACGATGTAAGCAATGTGCT 4560
QY 4561 gacggaacacctgcaaggtcgtgaacatctccagcaccatctgtgaagaacag 4620
|||||
Db 4561 GACGGCAACCTGCAAGGCTGTGTAACATCTCCACACCAATATGCTGAAGAAACAGTGC 4620
QY 4621 ccccaataagcgtgttctgaagcattctggaagcgaagagagagagagagagagag 4680
|||||

```

Db 4621 CCCAGATAGCGGCTGTTTCAGGACATCTGGACGAGCGCGAAGTGCAGTGTCTCCTG 4680
Oy 4681 aactacaacagaagaagataaagtcgtgagaacccaacccactactgcaatgaaac 4740
Db 4681 AACTACAAACAGAAAGGAGTAAGTGCCTGGAGAACCCAAACCCCTACTGCAATGAAGAAC 4740
Oy 4741 aatgagcggtgtgacgcgcgaatgcaatgacacagagaagacagcgctcctaacggaag 4800
Db 4741 AATGGCGGGTGTGACCGCGAGTAAATGACACCGAGGAAGACAGCGGCTCAACGGAAG 4800
Oy 4801 aaatcacatgagagtgactaagcccgactcctatccactcttcgacggaattttgc 4860
Db 4801 AAAATCACATGCGAGTACTAGACCCGACATCTATCCACTCTTCGACGGGATTTTGGC 4860
Oy 4861 tcgcgcctaatcttcctgagcattcctctctgctgactcctcactgctatcctgacagc 4920
Db 4861 TCGAGCTTAATTTCTCGGGCATCTCTCTCTGCTGATCTCATGCTGATCTGTCGACGC 4920
Oy 4921 ttcatctaataagatcagatg 4940
Db 4921 TTCACTTAATAGATGATGATG 4940

RESULT 3
PFAMT195
LOCUS 5917 bp mRNA linear INV 30-MAR-1995
DEFINITION Plasmodium falciparum mRNA for major surface antigens precursor (P195).
ACCESSION X02919.1 GI:9864
VERSION X02919.1
KEYWORDS antigen; direct repeat; signal peptide; surface antigen.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 5917)
AUTHORS Holder, A.A., Lockyer, M.J., Olink, K.G., Sandhu, J.S.,
Riveros-Moreno, V., Nicholls, S.C., Hillman, Y., Davey, L.S.,
Rizard, M.L.V., Schwarz, R.T. and Freeman, R.R.
TITLE Primary structure of the precursor to the three major surface
antigens of Plasmodium falciparum merozoites
JOURNAL Nature 317 (6034), 270-273 (1985)
MEDLINE 86014355
REFERENCE 2 (bases 1 to 5917)
AUTHORS Holder, A.A.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1991) Holder A.A., National Institute for Medical
Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
REMARK Update of published sequence
FEATURES
source Data kindly reviewed (06-MAR-1986) by R.R. Freeman.
location/Qualifiers
1..5917
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
418..474
/feature="putative"
418..537
/note="P195 precursor"
/codon_start=1
/protein_id="CAA26676.1"
/db_xref="GI:9865"
/translaton="MKIIFLCSPLEFIMTCVTHESYOELVKLEALDVAVLTVGS
LPQKEKVALNEGTSGAVTSTPGSKGVSAGSGSVASGVSAGSVASGVS
GGGNSRRTPSDNSDSADKSYADLKHRYRNYLLTKEKYPQLPFLTNHMLTCN
IHGFKYLDGEELNELLYKLNFYDLAKLNDVCAANDCOIPFNKIRANELDVLK
KIVGYRKLPLDNTKDNVGMEDYIKKKKTIENELIESKRTIDKNKATREEEK
KLYQADLSTYNNKOLEAHNLISVLEKRIIDLKNNINELDLKINELINPPANG
NTPNLTLDKRRKLEHEHEKLEIKAKTIFNIDSLFDPLELYLREKNNIDISAVY
ETKESTEPNEYPNQVYPLSYNDINNALINLSFGDLINPDYTKESKNIYDNEK
KFINIEKIKIEKKIESDKSYEDSKSLNDITKEKFLNKLINYSKNNINIDLK
FKNMGKRYSYVEKLTHTHTPASYESKSKNLKLTETKALYMDYSIRNVYVEKLY
YKNLISKEIEITLVENIKKDEQLFEKRTITDEKPRDKILIEVSIVYQOKVL
MKRIDELKTQTLILKNVELKHNHVPNSYOEKNKEPYLILVLEKLEIDSKLVKPKVE

```

```

SLINEKKNIKTEGOSDNSEPTGEITTCGATTKPGQAGSALEGDSVQAQAOEOKOA
OPPVAVPPEAKAOPVPPAPVNNKTNDYLEKLYEFENLSTYCHKYLIVSHST
MNEKILIOVKITREESKLSGCDPDLIFNIONNIIPWYMFPSLNSLSOLMEIYE
KEMYNITKIKDNDRKINLLEAKVSTSVKTLSSSMOPLSLTPOKPREYSANDPIS
HSTLNNLSLKFENILSLGKNNITPOLIOKSESENTERTLKOSDIFNEDSPFNRYK
SKADINDSLNDESKRRLLEEDINKLAKTLOLSDLYKRYLKLERLPDKRYGVKYM
QIKKLTLLKEQLESKLNNPKHVLONFVFNKKKEAEIETENLWKILKHY
KGLVYNGESSPLKTLSESIOTEDVYASLENKVLSEKLEKLDNLNLEKLSYL
SSGHHILIAELKRYIKKNYNTSPNSPNSPDVNNALNLEKLYPECDGADNVYSEGS
DYLESOPKRPASTHVGESNTITTSNVNDEVDVITVIFGESEDDYDLOAVYTG
EAVPVSVIDNLSITENEVLYLTKPLAGYRSLLKOLENNVMFENVAVDILSRN
KRENFKNVLESDLPYKDLTSSNVYNDPYKFLNKKERDFTLSYNYIKOSIDPINF
ANDVLYKLLSEKYSKSDLSIKKYINDKOGENKYPFLNNIETLYKNDKIDLYE
HLEKAVLNTYERKSNVEKIKELNTYLTODKADLDRKNNNFVGIADLSTDYNNHL
LTKFLSTGAFVFNILAKTVLSNLDGNLQMLANTSOHCVCVKOCGONGCPRHIDERE
CKCLINRKOGEDKVEPNPTCENNNGCCADAKCTBEDSGSNKKTTCCTRPDSIP
LFDGIFC5SSNFIIGISPLILMLILVSFI
475..5334
/product="region coding for 83k, 42k and 19k major surfa
antigens (aa 1-1628)"
631..728
repeat_region
/misc_feature
/note="alternating repeated SVA-SGG (aa 72-104)"
763..765
/misc_feature
/note="pot. glycosylation site"
1219..1221
/misc_feature
/note="pot. glycosylation site"
2707..2709
/misc_feature
/note="pot. glycosylation site"
2719..2721
/misc_feature
/note="pot. glycosylation site"
2764..2766
/misc_feature
/note="pot. glycosylation site"
2947..2949
/misc_feature
/note="pot. glycosylation site"
3175..3177
/misc_feature
/note="pot. glycosylation site"
3307..3309
/misc_feature
/note="pot. glycosylation site"
3589..3591
/misc_feature
/note="pot. glycosylation site"
3910..3912
/misc_feature
/note="pot. glycosylation site"
3937..3939
/misc_feature
/note="pot. glycosylation site"
4750..4752
/misc_feature
/note="pot. glycosylation site"
BASE COUNT 2644 a 636 c 727 g 1910 t
ORIGIN
Query Match 57.0%; Score 2816.4; DB 3; Length 5917;
Best Local Similarity 73.2%; Pred. No. 0;
Matches 3609; Conservative 0; Mismatches 1321; Indels 0; Gaps 0;
Oy 10 atgaatacatttctctcgtgtcatttctgtttttatcatcaatacagtgctg 69
Db 418 ATGAAGATCATATCTCTTTATGTCATTTCTTTTATTAATAACAAATGTGTA 477
Oy 70 acccaagatcctatcaggaagcgtgttaagaacgtgaagcgtttggaagatgcgctct 129
Db 478 ACACATGAAGCTTTATCAAGACTGTCAAAAAGACAGCTTTAAGAAAGGATGTTG 537
Oy 130 accgatacagccgtgttcagaagaagtgtcgaatagaaggaagcagtgacag 189
Db 538 ACAGTTATATTATTTCAAGAAATAATGATTAATAATGAAGAAACAAGTGAACA 597
Oy 190 gccgttacacacagacaccccggttctaaaggtctgtgctagcagtgctcggtgg 249
Db 598 GCTGTACAACTAGACACCTGTTCAAAAGGTTACAGTTCCTTCAGGTTCAGGTGAC 657
Oy 250 tctgtgctcttggtgggtctcgctgcctccggcgagcagcgatgataagtggtcagtg 309
Db 658 TCAGTGTCTCAGGTGCTCAGGTGCTCAGGTGCTCAGGTGCTCAGGTGCTCAGGT 717

```


QY 310 gcaagcgcggttcgcggagacagtcgaagaaccaatccatctgacaactctagcgtatcc 369
DB 718 GCTTCAGGTGGTTCCAGGTAATTCAGACGTAACAATCCTTCAGATATATTCAGTGAATTC 777
QY 370 gaagcgaagtctcaagccggaacctcaagcagctgagaactatctctcaactcaag 429
DB 778 GATGCTAAATCTTAATGCTGATTTAAACACAGCATGTAAGAAATTAATTAATTAATTA 837
QY 430 gaagcgaagtctcaagccggaacctcaagcagctgagaactatctctcaactctgatac 489
DB 838 GAACCTCAATATCTCACTCTTGAATTAACATATCATATGTTAACTTTGTGTGATAT 897
QY 490 attcatgcttcaaatatctgattgacggttgaagagatcaatgaactctgtacaag 549
DB 898 ATTCATGCTTCAATATTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 957
QY 550 ttgaattctactgactgactgctcaagggccaactgaatgagcttggcgcaactgactat 609
DB 958 TTAACCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1017
QY 610 tgcgaattccatccatctgaattgaaatcgaagccagctgagctgattgagaagtgtg 669
DB 1018 TGTCAATATACCTTTCAATCTTAATTTCTGCAATATGATTAAGCTACTTTAAANACTT 1077
QY 670 gtcttcgatatcgcaagcctctcgacaacatcaaggaacatgtgggaaagatlgagaat 729
DB 1078 GTCCTGATATATGAAACCATTAAGCAATATTAAGATATATGAGAAATATGGAAGAT 1137
QY 730 tatataaagaataagaagaccatcgagaacataatcaagagctgactcgagaatccaa 789
DB 1138 TACATTTAAAAATATAAAAAACCATTAAGAAATATTAATTAATTAATTAATTAATTAAG 1197
QY 790 aagaccatagacaataaagaatgcaaccaaggaagaagaagaagtgtgtacag 849
DB 1198 AAACATATGATTAATTAATTAAGATGCACTTAAGAAAGAAAGAAAGAAAGAAATATATCCAA 1257
QY 850 gccacatcgacccgtccatctataaacaacagcttgaagaagccataacccatcagc 909
DB 1258 GCTCATATATGATCTTCTTATTTACATTAACATTAAGAAAGCAATATTAATTAATTAAGC 1317
QY 910 gtactgagaagaagcagtagacacccctcaagaagaatgaataatgaagaagctgcagc 969
DB 1318 GTTTTGAAGAAACGTAATGACCTTTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1377
QY 970 aagattaatgaataatgaataatccctcgccagccaactctgagacaccccttaacagctg 1029
DB 1378 AAGATTAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1437
QY 1030 ctgagcaagaacaagaagatagagagacagagaagaagatcaagaagatcgccaaacc 1089
DB 1438 CTTGATTAAGAAACGAAAGAAATCGAGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1497
QY 1090 attaatgtcaaatct 1149
DB 1498 ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1557
QY 1150 gagaagaataagaataatagacatctcgccaaagctcgagacaagaagaatcaacagcaact 1209
DB 1558 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1617
QY 1210 aatgaataatccaaatgtgtgacgtacccctgtctctataagatatcaacaagctctc 1269
DB 1618 AATGAATATCAAAATGAGATGATCTTATCTCTTGTATATTAAGATATTAATTAATTAAT 1677
QY 1270 aacgagctcaatagctgtgtgacgtgattgaaccccttgattatatacgaagaacccctc 1329
DB 1678 AATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1737
QY 1330 aagaatctatacagaacaatagagaaagatttatacagaagaatcaagagagatc 1389
DB 1738 AAAAAATATATATCTGATATATGAAGAAAAAATTCATTAATTAATTAATTAATTAAGAA 1797
QY 1390 aaaaatcgagaagaagaataatgtgagatlgacaagaagaatgacgaagccgcaagaat 1449
DB 1798 AAAATGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1857
QY 1450 cttaacgatatctcaagaagatlgaaaagctgtctgaagagatctatgtatccaatctc 1509
DB 1858 TTAATGATATTAACAAAGAAATATGAAGAAATTAATTAATTAATTAATTAATTAATTA 1917
QY 1510 aacaataacatcgacccctcgacccaactcgagaagaatgaggaagaacgtatctctcaaa 1569
DB 1918 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1977
QY 1570 gtgagaagaactgacacacaataatcccttgcatctctatagagaatctgaagcaatctc 1629
DB 1978 GTTGAGAAACTTTCACACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2037
QY 1630 gagaagctccacaagcctcttaagatatgaggagactatctctcggaacatgtgtgtg 1689
DB 2038 GAAAGCTTAACAAAGCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2097
QY 1690 gagaagaacataagatatacaagaatctcatatagtaagaatcgaaaaagagatcgagag 1749
DB 2098 GAAAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2157
QY 1750 ctgtgtgagaacataagaagaatgaagaacagctgtcttgagaagaatatacaaaagc 1809
DB 2158 TTAGTTGAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2217
QY 1810 gaaaataaaccagaatgagaagatcccgagagctctcgatattgttaaatccaaagtcag 1869
DB 2218 GAAATTAACCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2277
QY 1870 aaggtgctccctatgacaagaatgtgaatctcaagaagaactcaactctctgaagaac 1929
DB 2278 AAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2337
QY 1930 gtgagatlaaacaataatatactgtgcgaagatgataagcagagaagaataagcagaa 1989
DB 2338 GTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2397
QY 1990 ccaactactactctgactcgaagaagaatgagaagaagaactgaagaagtgttcatgccc 2049
DB 2398 CCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2457
QY 2050 gtccgagccctgtatcaagagaagaagaagaacattaaactgaagaagatcagaataac 2109
DB 2458 GTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2517
QY 2110 tccgagccttccacaagaagagagataacccgagcaggtctcaaccaaagcccgagaac 2169
DB 2518 TCGGAACCAATCAACCGAAGAGAAATTAACAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2577
QY 2170 gccggttcagcctcggaagcgaatagctgtgcaagctcaagcacaagagcagaagcaga 2229
DB 2578 GCAGGATCTGCTTTAAGAGAGATTCAGTACAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2637
QY 2230 cagcctcaggtgtgcagtgccgttccagaggtctaaagctcaaggtgctcaacacacagct 2289
DB 2638 CAACCAACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2697
QY 2290 cctgtgataatacaagaacgagaatgtcgagcaaacctgtgactaccttgagaagctctatg 2349
DB 2698 CCAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2757
QY 2350 ttccgtgaataactctcaagctctgcaacaatatatactctgtctctcaagcctatgaac 2409
DB 2758 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2817
QY 2410 gagaagatcttlaaacaagatacaagaatacaagaagaagagagaatgataatctctctgt 2469
DB 2818 GAAAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2877
QY 2470 gatccactggaacctgtctgtatcaatataccagaacaacatcccggtatgtatctatgtc 2529

QY	4690	caagaaggaagaaatg	gctgctgctggaagacc	caaacccctactctgcaatg	aaacaatg	cgcg	4749
Db	5098	CAGAAGGTGATTAATG	TGTTGA	AAATATCCAAATCTCTACTTGT	TAACGA	AAATATATGTGGA	5157
QY	4750	tgtagccgcgaatg	ctcaaatgcaacggaag	gaagacagcgctc	ctaacggaaga	aaatctaca	4809
Db	5158	TGTGATGCAAGATGCCAA	ATGTCACGAAGAAGTT	AGAGTACGAC	AGCAAGGA	AAATATACCA	5217
QY	4810	ttcgagatgtaactaag	cccgcaactctatcaact	cttcgacgaggaat	tttctgctcagctc	4869	
Db	5218	TGTGATGCTACTAAAC	CGATTCGATCTTATCCATTTTCGATGATTTTTC	GTGCGATTCCTCT	5277		
QY	4870	aattccggcgcatctc	ctctctgctgtagctccat	gcgcgcatccctgtaacgtc	taacgtc	taac	4929
Db	5278	AACCTCTTAGGAAATAT	CATCTTTATTATTA	TACTCATGTTAATATTA	TACAGCTTCA	TTTAA	5337
QY	4930	tagatcagtg	4939				
Db	5338	AAAATGTAGG	5347				
RESULT	4	PF0P195A	7038 bp	mRNA	linear	INV 03-APR-1995	
LOCUS	PF0P195A						
DEFINITION		Plasmodium falciparum mRNA for major merozoite surface antigen					
ACCESSION		gpi95					
VERSION		X15063.1	GI:9896				
KEYWORDS		antigen; glycoprotein; gpi95 gene; surface antigen.					
SOURCE		malaria parasite P. falciparum.					
ORGANISM		Plasmodium falciparum					
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
AUTHORS		1 (bases 1 to 7038)					
TITLE		Myler, P.J.					
JOURNAL		Direct Submission					
AUTHORS		Submitted (11-APR-1989) Myler P.J., Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle WA 98109-1651, U S A					
TITLE		2 (bases 1 to 7038)					
JOURNAL		Myler, P.J.					
MEDLINE		Nucleotide and deduced amino acid sequence of the gpi95 (MSA-1) gene from Plasmodium falciparum Palo Alto PLF-3/Bil					
FEATURES		Nucleic Acids Res. 17 (13), 5401 (1989)					
SOURCE		89345116					
		Location/Qualifiers					
		1. 7038					
		/organism="Plasmodium falciparum"					
		/strain="Palo Alto PLF-3/Bil"					
		/db_xref="taxon:5833"					
		/dev_stage="trophozoite/schizont."					
		1592..1648					
		1592..16511					
		/codon_start=1					
		/product="gpi95 surface antigen preprotein"					
		/protein_id="CA33163.1"					
		/db_xref="gi:9897"					
		/translation="MKILIFPSCLPFIINTOCVTHESYQELVYKLEALDEAVLTGVS					
		LFOEKRWYLVNGTSGTAVTTSPTSPKSGVAGGGGSAVSGSVASGGSVASGGS					
		GGGSSNRRTNDSNDSODAKSVADLKHRYNYLTITKELYLPOLFDLTNMLTCDN					
		IHGKRYLIDGEEINELLYKLNFFDLIRALINVCADYQOIFPNLTIRANEIDVLYK					
		KLVGKRPDLKEDINQNGKMDYIKNKKTLEINNELIBSKRTIDKNKNATKEBEKK					
		KLVOAYDLSTYNQLEBANHLISVLEKRIDTLKKNNEIKELDINIKNPPANSQ					
		NTPLDLDKNKIEBEHEKEIKETIKFNIDSLFTPDLELEYLRKKNIDISAKVY					
		ETKSTEPNEVPNYTPPLSTNDINNALNELNSGDLINPDYIKREPSKNITDNERK					
		KFIVIEKIKITERRKLESDDKSTEDRSKSLNDITKVEKLTNLEYSKFNENNIDTLC					
		PEKMKRKYKRVKLEHTHNPASTYENSKNMLERLTAKLYME					

[illegible]

Qy	790	aagccatagacaaaataaagaatlgcaacccaagsggaaagaagaagtgtaccag	849
Db	2372	AAAACAATGTATATAAATAATGAATGCACTTAAGAAGAGAAAAAAATTAATCCAA	2431
Qy	850	gccagtagcaccgtgccatctaaacaacgcttgaaagcccatcaactcatcag	909
Db	2432	GCCTAATATGATCTTCTCTATTATTCATTAACATTAAGAAAGACATTAATTATAAGC	2491
Qy	910	gtactgagaagcgcatagacaccctcaagaagaatgaaalatacaagaacgtctcgac	969
Db	2492	GTTTTCAGAAAAAGCGTATTGACACTTTTAAAAAAAATGAAAACATTTAAGCAATTCTGAT	2551
Qy	970	aagatlaaigaattiaagaatccctccgacgcacactctgggaacacccctcaacgctg	1029
Db	2552	AAGTAAATGAAATTAATAAATTCGCCACGGCCAAATCTTGGAATATACCAATATCTCTC	2611
Qy	1030	ctggacaagacaagaagatagagagacacgsgaaagatcaeaagatcgcgcaaac	1089
Db	2612	CTTGATATAGAACAAAAATTCGAGAACACGAAAAAGAAATTAAGAAATTCGCAAAACT	2671
Qy	1090	atgaagtcagacagtagatctcctcttactgtgcccttgagctcgagtagtactactaga	1149
Db	2672	ATTAAATTTAATGTGATAGCTTATTACTGATCCCTGGAATTGAAATCTATTTAAGA	2731
Qy	1150	gagaagaataagatalagacatctccgcgcaagctcgagacaagaagaaatcaacggaact	1209
Db	2732	GAAAAAAATTAATAATATGTGATTAATAGTCGAAGAAGTTGAACAAAGATCAATGAAACC	2791
Qy	1210	aatgaatatacccaatggtgtgtagagtaacctctgcttataeaagatatacaaacgctctc	1269
Db	2792	AATGAATATCCAAATGGAGTACTTATCCCTTGTCATATAAAGATTAACGAATTAACGAATGCTTTA	2851
Qy	1270	aacgagctcaatagctctggttagactgattcaacccctcgattatagcaagaacccctc	1329
Db	2852	AATGAATTAATCTTTTGGTGATTTAATTAATCCATTTGATTATACAAAGAACCAAGT	2911
Qy	1330	aagaaatctacaacagaatlgagagaagaagttatatacaagaaatcaagagaagaatc	1389
Db	2912	AAAAACATATATCTGATATGAAAGAAAAAAATTCATTAATGAATTAACGAAAAAATT	2971
Qy	1390	aaactlgagaagaagaanaattlgagltgacgaagaagttacgaagaccgcgaagaaagt	1449
Db	2972	AAATATGAAAAAATAAATTTGAATCTGATTAATAAATCTTCAGGAAGACAGATCTTAAGCT	3031
Qy	1450	ctaaacgatalctacctaagaagta.tgaanaagctgcgtgaacgagatctagatccaatc	1509
Db	3032	TTTAATGATATACAAAAAGAAATATGAAAAATTACTTAATGAATTAATGATAGCAAAATTC	3091
Qy	1510	aacaaataactgaacctgacccaactctcgagaanaatgatatgggaaaaacgctactctaca	1569
Db	3092	AATATATATATTAATATTAACTTAATTTTCGAAAAAATGATGGTAAAGATATCTCATATAAA	3151
Qy	1570	gtgagaagaactgacacacctaatactactgcatcciatagaatctcaagcataact	1629
Db	3152	GTTGAGAACTTACACACCATATACTTTTGCATCTCTAGAAATTTCTTAACATATATCTT	3211
Qy	1630	gagaagctcaaccaagctccttaaglatatagtagagacatctctcgtcggaacatgtgtg	1689
Db	3212	GAATAAGTTAAACAAAGCTCTTAATATATATGGAAGATTAATCTTTTAAGCAATATAGTAGT	3271
Qy	1690	gagaagaagaactaagaatcatcaagaatctcataagaatagatcgaaacagaatcgagaag	1749
Db	3272	GAATAAGATTTAAATATTTTATAAATTTTAATTAAGCAAAATTAAGAAATTAAGCAATTAAGCA	3331
Qy	1750	cttgttgagacatlaagaagaatlgagaacagctgtgttgagaaagaagatcatcaacaagac	1809
Db	3332	TTAGCTTGAATAATATTAATAAAGATGAAGACAGCTTTTGAATAAATAAATTAAGTAAGAC	3391
Qy	1810	gaataataaccagatagagaagatcccgagagctctccgatatgtttaaagtcgaagtgcag	1869
Db	3392	GAATAATTAACCAATGAAAAAATTTTATAGAAAGATATCTGCATATTTTAAGTCAAGTTCA	3451

QY	1870	aagygctccctcctatgaaacaagattgtaagtaactcaagaagctcaactcctcgtgaagac	1929
Db	3452	AAAGTTTTATTAAATCAACAAAATTGACAAATTTAAAAAAAGACTCAATGATTTTAAAAAT	3511
QY	1930	gtgagagcttaaaacatacatatcatgtgcgcgatagtctaaagcaggaagataagcagaa	1989
Db	3512	GTAGAAATTAAACATAAATATATACATGTTCCCAATTTCTTACAAACAGAAATAATACCAAGAA	3571
QY	1990	ccatactacatcatctgctcctcaagaagaaga taagaacaactgaaagtgctcctgcccana	2049
Db	3572	CCTTATTTATTTAAATGTGTGAAAAAAGAAATTTGATAATTTAAAAAGTGTGATGCTTAAG	3631
QY	2050	gtcgaagagctgatactcaacgaagaagaagaacattaaactgaagaagcagtcagaatac	2109
Db	3632	GTAGAAATCATGTGATTAAGAGAAAAAAAACATATAAACAGAAAGGTCATCGGATTAAT	3691
QY	2110	tccgagccttccacaagaagaagagataaaccgaagagctaccacaagaagcccgagcaacag	2169
Db	3692	TGCGAACCATCAACCCGAAGAGAAATTAACAGGACACAGCACTTACAAAACCTGGACAAACA	3751
QY	2170	gccggtttcagctctcgaagcgatagcgtgcgaagctcaagcacagaagcagaagcagca	2229
Db	3752	GCAGATGTCGTTTGAAGAGAGATTCAGTACAGACACAGACAAAGCAAAACAAAGCA	3811
QY	2230	cagctcagtgccagtgccggttccagaagctaaagctcaagtgcctacacacagct	2289
Db	3812	CAACCAACAGTACCGAGTACAGATACCAACAAACAAAGCAAGTCCCAACCAACAGCA	3871
QY	2290	cctgtgaataacaagaccgagaatgtcgaacaaactggaactccttgagaagctctatgag	2349
Db	3872	CCAGTAAATTAATTAACACTGAAAATGTTCCAAATTTAGATTTCCTTGAAAAATTTATAGAA	3931
QY	2350	ttccctgaaatacctcctacatctgcacacaatafatcctcgtctctcacaagcactatgaac	2409
Db	3932	TTTTAAATACTCTCATATATATATGTCACAAAATATATTTGGTTCCACTCAATCATATGAC	3991
QY	2410	gagaagattcttaaacagtaacaagataaacaagaagaagagagtaactgtcctctgt	2469
Db	3992	GAAAGAAATTAATTAACAAATTAATAATTTCAAAAGGGAGAAACCAAAATTAAGTTCAATGT	4051
QY	2470	gattcaacggagcgtgttcaatatccagaagaacatctccgctatgatctcatgtc	2529
Db	4052	GATCCATTAGATTAATTTGTTAAATATACAAAATACATACCTGTAAATGATTTCTATGTTT	4111
QY	2530	gatagcctccaacatctcctctcctcaactgtctcgaatgataatgagaaggaagatgctc	2589
Db	4112	GATACTTAATACAAATAGTTTATACAACTATTTTATGGAATTTTGAAGAAAAAGAAATGTT	4171
QY	2590	tgcacactgtataaaactcaagaacaacgaagaagatlaagaacctcttggaggaaactaag	2649
Db	4172	TGTAAATTTATTAACCTTAAGAGATATGACAAAATTTAAAAATTTATTAAGAGAGCGAAA	4231
QY	2650	aaggtctcaccctctgttaaaactctctctccagcttccatgagaacaactgtctctaca	2709
Db	4232	AAAGATATCCACATCGTGAATAAACTCTTTCAAGTTCATCAATGACCAACTTATTCATTAACA	4291
QY	2710	cctcaagaacaagcccgaaagttagcgctaaagcagacacactctcaactcgaccaacttaat	2769
Db	4292	CCTGAGATTAACCCGGAAGTAAAGTCAATATGATACATCACTTCTCAAAATTTTGAT	4351
QY	2770	aactcaactgaanaactgtttagaacaatctcgtctcctcggcaagaataaagaacatctacca	2829
Db	4352	AATAGTTTAAATATTTATTTGAAAACCATATTTAGTCTGTGAAAAAACAACAAAATATATACCA	4411
QY	2830	gaactatcttgacagaaatctgcgagaactctcaagagaagatactcgaagaagaagcgac	2889
Db	4412	GAATTAATATAGTCAAAAAAGTAGTGAACACTTTTATGAAAAAATTTAAAGATATGAT	4471
QY	2890	acattcttaacaggaagcttcaactaactgtggaacttaagccgagataatcaactc	2949
Db	4472	ACATTTTATTAATGAATCTTTTACAAATTTTGTAAAAATCTTAAGCTGATGATATTAATTC	4531
QY	2950	cttaacagctgaatctaaacgtlaagaagcttgaagaagagacatcaataagctgtgaagaagaca	3009

Db	4532	TTGAATGATGATCAAAAAGGAATTTAGAAAGATATTATTAATTTAAAAAAACT	4591
Qy	3010	ctcgaactcgagcttcgacctcgatacaagttacaaactcgaactcggagagctctcgac	3068
Db	4592	TTAAAGTTATCATTTGGATTTTATTTATTAATAATAATTTAAAAATTGAAAAGATTATTTAT	4651
Qy	3070	aagaagaagacgctcgacaagttataagatgcagatcaagaacttgactctgcacaagag	3129
Db	4652	AAAAAGAAACAGTTGGTAAATATTTAAATGCCAATTAAAAAACTTACTTTATTAAGAA	4711
Qy	3130	caagcttgaaagcaaaactcaactcgaacaatccgaacacgctactcgaagaacttcca	3189
Db	4712	CAATTAGATCACAAAATTTGAATTTCACTTAATTAACCCAAACACATGTATTACAAAACCTTCT	4771
Qy	3190	gttcttcctcaagaagaaggaagccgagatccgagacagagaagaacactctgagac	3249
Db	4772	GTTTCTCTTAACAAAAAAGAAAGCTGGAATTCGCAAACTGAAACACATTACAAAAC	4831
Qy	3250	accaagaattctctccaacaactcaaaagcctcgtaagttatataatgacgagctctc	3309
Db	4832	ACAAAAATATTTAGAAACATTTTAAAGGACTTTTAAATATTTAATATGCGAATCAGCT	4891
Qy	3310	ccctctgaagactctctccggagagagacatccgaaccgaggtataactagccagcctcgag	3368
Db	4892	CCATTTAAAACTTTTAAGTGAAGATCAATTTCAACGAAAGATTAATTAATCCAGTTTAA	4951
Qy	3370	aactccaagtcctcgcttaagctccgaagcaagctbaagagaacactcgaaactcgagag	3429
Db	4952	AACTTTAAAGATTTAAGTAAATTAGAAAGGAATTTAAAGATTAATTTAAATTAAAG	5011
Qy	3430	aagaagctcaagctacccctcttaagcggactcgatcaactcgatcgcgcgagctcaaggaagtc	3488
Db	5012	AAAAAATATCATATCTTATCACTGATTAACATCACTTAAATTCGTAATTAAGAAAGTA	5071
Qy	3490	altaagaagaagaactacacccggacaatagccgaagcggaatataataagacgtgaataac	3549
Db	5072	ATTAATAATAATAATTTATACAGTAAATTCTCCAGAGTGAATAATATAGGATTTAAACAT	5131
Qy	3550	gcactcgaaacttacaagaagttccctcgcttgaaagaaacagatgctgcacgtgtagtctc	3609
Db	5132	GCATTAGATCTTACAAAAAATTTCTCCGAAAGGAACAGATGTGGCAACACTTTAAGT	5191
Qy	3610	gaactcgctccgcacacatcggaagcagttccaaactcaagaagccgcgcactcactatgct	3669
Db	5192	GAAAGTGGATCCGACACATTAGAACAAAGTCACACCAAAAGAACGACATCAACTATGTA	5251
Qy	3670	ggagccgaftccaaatacaattaccacaatcctcagaacgctgcagatgtagtgcatagcctc	3729
Db	5252	GGACACAGAGCTTAACACAAATTAACACATCAACAAAATGTGATGATGAATAGATGACGTA	5311
Qy	3730	atacttgctcattctcgcgcgagagcgagagactacgaatgactcgcgcgcgcagtgctc	3789
Db	5312	ATCATATGACTATATTTTGGAGAAATCGAAGAAAGATTATGATGATTTAGACAACTATGTA	5371
Qy	3790	accggtgaagcgtgcactccctccgcgtatgtataaactctgccaaatcgaagaacga	3849
Db	5372	ACAGAGAGACAGTAACCTCCTTCCTCGTATTTGATTAACATCTTCTTAATTTAAAAATGA	5431
Qy	3850	taagaagtgcctacttgaaacactctgcaagcgctctatagtlctctcaagaacagctg	3909
Db	5432	TATAGAGTTTATTTATTTAAAACTTTAGAGCGGTTTATAGAAAGTTTAAAAAACATTA	5491
Qy	3910	gagatatagctgtagactcttcaatctgtaacagtggaagaacactctgaacagccgcttat	3969
Db	5492	GAAATATACGTTATGACATTTAATGTTAATGTTAAAGATATTTTAAATTCACGATTATAT	5551
Qy	3970	aagaagagaaattccaagaacgctctggagagcgacttgattccctataaagaactcgacc	4029
Db	5552	AAAGCTGAATTTTCAAAAATGTTTTAGATCAAGATCAATTAATCAATTAAGATTAAACA	5611
Qy	4030	tcctctactaagcttgctcaaggaacccatacaagttccctcaataaagaagaagataaa	4089

D	5612	TCAAGTATTTGTTGTCCAAAGATCCATTAATAATTCTTAAATGAAGAAAAAAGACATAAA	5671
OY	4090	tttcgtctagtattcaactatatacgaagccctccgcacccgatatcaatttgcctat	4149
D	5672	TTCTTAACCGACTTAATAATTATTAAGAATTCAAATAGATACGGATTAATAATTTTGCAAT	5731
OY	4150	gatctgcgggagattacaagaatccctgagcgaaaaatacaagtcgcaccttgactctat	4209
D	5732	GATGTTCTTGATATTTATAAATATTATTCGCCAAAATATTAATCAGATTTTAGATTCAAAT	5791
OY	4210	aaaagtatatcaacgataaagcaaggcgagaaatgaataatctgccttccctgaataac	4269
D	5792	AAAAATATATCACGCAGCAAAACAAGGTAAAAATGGAATACCTTCCTTTTTTAACAT	5851
OY	4270	atcgaaccctctcaaacagatgtagcaccaaatcgacccttcctgatatccaactggag	4329
D	5852	ATTGAGACCTTATATTAACACGTTATATATAAATTTGATTTATTGTATATTCATTAGAA	5911
OY	4330	gccaaagtcctcaactactacttaacgagaagacaatlgtagaagttaaatacaaggagctg	4389
D	5912	GCAAAAGTTCTTAATTTATACATATATGAGAAATCAACACGTAGAAAGTTAAATAAAGACTT	5971
OY	4390	aactaaccccaaaaataatccaagaacagatggcagatttcaagaanaataaacattgttc	4449
D	5972	AATTACTTAATAAACATTCAGACCAAAATGGCAGATTTTAAAAAATAATACATTTTCGT	6031
OY	4450	ggaattcgagaacctgtctacacgatataaacaccaaaaatctctcgaaccaagtlctgtcc	4509
D	6032	GGAATTCGTGATTTATATCAACAGATTAATACCATATTACTTAATGACAAAGTTCCTTAGT	6091
OY	4510	actgcatgagtggttcogaaaacctgcgcaaaacagtcgcgaacatctgtcgaaggcaac	4569
D	6092	ACAGGATAGGTTTTTGAAATCTTGCTAATAACCGTTTATCTAATTTACTTGATGGAAC	6151
OY	4570	ctcgaaaggcagctcgaaacatctcccagaaccaatlgcgtgaagaacaagtgccccagat	4629
D	6152	TTGCAAGGATAGTTAAACATTTACCAACACCACCAATGCGTAAAAAACAATGTCCAAAT	6211
OY	4630	agcgctgtttccaggaatcttgagcagagcggaaggtgcaagtgctcctcgaactacaaa	4689
D	6212	TCTGGATTTTTCAGCATTTAGATGAAGAAGAGATTAATGTTTTTAATATACAAA	6271
OY	4690	caagaaggagataaagtgctgtgagaagaccaaaacctactcgaaatgaaanaacatgycgg	4749
D	6272	CAGAAGGTTGATTAATATGTTGGAANAATGCCAATCTACTTGTGAACGAAATATGTGTGA	6331
OY	4750	tgtgacgcgcgatgtctaataatgcacgcggaagaacagcggtcctaacygaagaataatcaa	4809
D	6332	TGTGATGCAAGATGCCCMAATGTACCCMAAGATTCAGGTAGCAGCAAGGAAGAAATTCACA	6391
OY	4810	tgcgagtgataaagcccgacatccatcacactctcgcgcggagattttgtcccaagcct	4869
D	6392	TGTGATTTACTAAACCTTGATTTCTTATCCACTTTTTCGATGGATTTTCTGCAAGTTCCCT	6451
OY	4870	aatttcccggaatccctcctcctgctgtaatcctcatgctgtaactctgtaactctaa	4929
D	6452	AACCTCTTAGGAATATCATCTTATTATTAATACATGTAATATTAACAGTTCATTAA	6511
OY	4930	tagatcatg 4939	
D	6512	AAAATGTAGG 6521	
RESULT	5		
A92450			
LOCUS	A92450	4920 bp	DNA linear PAT 22-JAN-2000
DEFINITION	Sequence 1 from Patent W09814583.		
ACCESSION	A92450		
VERSION	A92450.1 GI:6741180		
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unidentified		
	unclassified.		

RESULT	5		
A92450			
LOCUS	A92450	4920 bp	DNA
DEFINITION	Sequence 1 from Patent WO9814583.		Linear
ACCESSION	A92450		
VERSION	A92450.1	GI:6741180	
KEYWORDS	.		
SOURCE	unidentified.		
ORGANISM	unidentified		
	unclassified.		

QY 1930 gtagagtaaaacataacatagtcggaagatgataagcaggaagaa 1989
 DB 1921 GTGAAATTAACATATATATACATGTTCCCAATTCACAAACAGAAATTAAGCAAGAA 1980
 QY 1990 ccaactactactcactgactcaagaagaagatagacaactgaaagtgctcatgcccacaa 2049
 DB 1981 CCTTATTTATTTATTTGTTGTAAGAAAGAAATGATTAATTAAGATGCTCATGCTTAAG 2040
 QY 2050 gtccgaagccctgcaacgaagaagaagaacatlaaactgaaagcagtcagataac 2109
 DB 2041 GTGAAATCATTTGATTAATGAAGAAAAAATAACATAAAACAGAAAGTCAATCGATTAAT 2100
 QY 2110 tccgaagcttccacagaagaagagataacccgagacagctccaccaaagcccggagcaag 2169
 DB 2101 TCGGAACCATCAACCCGAAAGAGAAATTAACAGCAACCAACTCAAAAACCTGGCAACAA 2160
 QY 2170 ggcggttcaagctcgcgaagcgatagcggtgcaagctcaagcacaagaagcgaagcagca 2229
 DB 2161 GCAGGATCTGCTTTAGAGAGAGATTCAGTCAAGCAACACACACAGAACAAAAACAAAGCA 2220
 QY 2230 cagcctccagtgccagtgcccggtccacagagctaaagctcaagtgccctacacacacagct 2289
 DB 2221 CAACCAACCACTACACAGTACAGTACAGCAAGCAAAAGCAAGTCCCAACACCAACAGCA 2280
 QY 2290 cctgtgaataacagaagccggaatgltcagaacactgagctacacttgagaagctctatgag 2349
 DB 2281 CCAAGTAATATATAAATGAAATGTTTCCAAATTAAGATTATCTGAAAAATTATATAGAA 2340
 QY 2350 ttccctgatacatcctcactacatcgcacaaatatactcctcctcctcacaagcactatgac 2409
 DB 2341 TTTTAAATATCTTCATATATATGTCACAAATATATTTTGGTTTCCACCTCAGCATATGAC 2400
 QY 2410 gagaagatcttaacacagtaacagataacgaagaagaagaglaaacctgtccctctgt 2469
 DB 2401 GAAAGATATTAATAAACATATTAATAATTAACAAAGCAGCAAAAGCAAAATTAAGTTCATGT 2460
 QY 2470 gaccactgagactgctggttcaatataccagaacacactccggtatgcttcatgtt 2529
 DB 2461 GATTCATTAAGCTTATGATTTATTAATACAAATTAACATACCTGTAAGTATCTATGTTT 2520
 QY 2530 gataagctcaacaactcctcctcctcactgltcactgagagatataagaaagagatgtc 2589
 DB 2521 GATAGCTTAACAATAGTTTATTCACAACATTTATGGAATTTATGAAAAAGAAATGCTT 2580
 QY 2590 tgcacactgtataaactcaagaacacagacagagatlaagaacctctgsgaagaaagtaag 2649
 DB 2581 TGTAAATTTATATTAACCTTAAGATTAATGACAAATTAATAATTTATTAAGAGAAAGCGAAA 2640
 QY 2650 aaggtcccaacctgtgtaaaactctctcctcagctccactgcaacacagctctctaca 2709
 DB 2641 AAAGTATCCACATCTGTAAAAAATCTTTCAAGTTTCATCAATGCAACCATTAATCATTTACA 2700
 QY 2710 cctcaagaacagcccggaagtgcggtcgaacgacagacacccctcactgcgaacacactaat 2769
 DB 2701 CCTCAGAGATTAACCCGAAGTAAGTGCAAAATGATGATACATTCACATTTTCAAAATTTGAAT 2760
 QY 2770 aactcactgaaactglttggaacatcctgtctcctgcggaagaatlaagaaacatctaccaa 2829
 DB 2761 AATAGTTTAAATTTATTTGAAAAATATGAGTCTTGAAAAAACAATAATATATACCAA 2820
 QY 2830 gaacttatggaagaagaatcgtccggaactctcagagaagaatctcgaagaagcagagac 2889
 DB 2821 GAATTAATAGCTCAAAAGTACTGTAAGAAATCTTTATGTAAGAAATTAATAAGATATGAT 2880
 QY 2890 acattctataacgagactcactcaactcgtgaaactcgaagcgaatgaatcaactc 2949
 DB 2881 ACATTTTATTAATGAATCTTTTACAAATTTTGTAAATCTTAAGCTGATGATTAATTTCA 2940
 QY 2950 cttaacgaatgaatctaaagcgaagaaagctgcgaagagagacatcaatlaagctgaagaagaca 3009
 DB 2941 TTGAATGATGATCAAAAGAAAGAAATTAAGAGAAATTAATTAATTAATAAATAAAT 3000
 QY 3010 ctgcgaactgagcttgcagccgtctacaacaagtlacaactgaaactgcgagagacactcttgac 3069

DB 3001 TTACAGTTATCATTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGAT 3060
 QY 3070 aagaagaagcagctgcgcaagtataagatgcagatcaagaagctgcctgcctcaagag 3129
 DB 3061 AAAAGAAACAGTGTGTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3120
 QY 3130 cagcttgaagaacaaactcaactcactgaaacatccgaagaacgtaactgcgaactctca 3189
 DB 3121 CAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3180
 QY 3190 gttgtcttcaagaagaagaagccggaatgcgcgagcagaagaacactcctgcggaac 3249
 DB 3181 GTTTCTTTTAACAAAAAAGAAAGCTGAATATGCAAGAACTGAACACATTAAGAAAC 3240
 QY 3250 accaagatctctcctcaacactcaagaagcctgcgtcaagtattataatgagcagctcct 3309
 DB 3241 ACAAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3300
 QY 3310 cctctgaagactcctcgcgagagagacatccagaacggagataactacgcgaactcgag 3369
 DB 3301 CCAATTAATAAATCTTAAGTGAAGATCAATTAACAAACAGAAATTAATTAATTAATTAATTAATTAAGAA 3360
 QY 3370 aacttcaagctcctgtcttaagctcgaagcgaagctgaagcgaacactgaaactgcggaag 3429
 DB 3361 AACTTTAAAGTATTAATTAATTAATTAAGAGAAATTAAGAGATTAATTAATTAATTAATTAAGAAAG 3420
 QY 3430 aagaagctcagactcctcctcctcctcagagctgcagactcagctgcgcgcagctcaagaagatc 3489
 DB 3421 AAAAATTAATCATCACTATCAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3480
 QY 3490 attaagaagaagaactacacccggcaatagcccaagcgaagaataatcacagcgtgataac 3549
 DB 3481 ATAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540
 QY 3550 gcaactggaactcctcaagaagactcctgcctgcgaagaagcagatgtgcacactgcgtgctc 3609
 DB 3541 GCATTAAGATCTTAACAAAAATTTCTCCCAAGAAAGAAACAGATGTTGACACAGTTGTAAGT 3600
 QY 3610 gaactggtccgacacactgcgagactcctcaactaagaagcctgcatctactcactgctc 3669
 DB 3601 GAAAGTGAATCCGACACATTAAGAAACAAAGTCAACCAAGAAACAGCATCACTCATGTA 3660
 QY 3670 gtagcgcagatcccaataacatccacatccgaagctgcgaagatgagatgcagatgcctc 3729
 DB 3661 GGAGCAGAGCTTAACACAAATTAACAAACATCAAAATGTCATGATGAAGTATGACGTA 3720
 QY 3730 atcatgtgctcactcctcgcgcgagagagagagactagaatgacactgcgagctgctc 3789
 DB 3721 ATCATTAATCACTTATTTTGGAGAAATCCGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 3780
 QY 3790 accgttgaagctgtcactcctcctcctgctgctgataacatctgtccaaatccgaagcgaag 3849
 DB 3781 ACAGGAACACAGTAATCTCTCCGTAAATGATTAACATTAATTTCTTAATAATTAAGAAATTAAGAA 3840
 QY 3850 taagaagctcctatactgaaacactgcgcaagcgtctatagctcctcgaagaacagctg 3909
 DB 3841 TATGAGGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3900
 QY 3910 gagaataagctgtgactcactcaatgcaacgltgaagacactctgcaacagccgcttaac 3969
 DB 3901 GAATAATTAACCTTATGACATTTAATGTTAATGTTAAGGATATTTTAAATTTACATTTTAAT 3960
 QY 3970 aagaagaagaactcctcaagaagcgtcttggaagagactgtatccctataaagaacgcctgac 4029
 DB 3961 AAACGTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4020
 QY 4030 tcccttaacagctgtcgaagcgaacatcaagctcctcaataaagaagaagagatataa 4089
 DB 4021 TCAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4080
 QY 4090 ttctgtcagttacactatccaaagactccatcgacccgatatcaatcttcgctaat 4149

Db	4081	TTTAAACAGAGTTAAATTTATATTAAGGATTCATACAGTATACGGATATATAATTTTGCATAT	4140
Oy	4150	gatcgtcgtgggcatlaccagaatccctgacggaataacaagtcgtacccttgactatc	4209
Db	4141	GATGTTCTTGATGATTTATTAATAATATATATTCGGAATAATATCAAGATTGTGATTCATAT	4200
Oy	4210	aaaaagtatcatcaagatataagaagcgagatgtgaaatatctgcacctctctgataaac	4269
Db	4201	AAAAAATATATCAACGACAAACACAGGTGAAAATGGAATATCTCCCTTTTAAACAT	4260
Oy	4270	atcgaagaacctgtacaagaacagltgacagcaaaaatcgacctctctgtaaccctggag	4329
Db	4261	ATTGAGACCTTATATATAAAACAGTTAATGATTAATAATTTGATTTATTTGTAATTCATTAGAA	4320
Oy	4330	gccaaagtcctcccaactataactlaccagaagaagcaatltgaggaaagttaataatcaagaagctg	4389
Db	4321	GCAAAAGTTTCAAAATATATACATATGACAAATCAACCGTAGAAGTAAATATAAAACACTT	4380
Oy	4390	aactaccctcaaaaacatcccaagaagaagltggcagaattccaagaaaaataacatctgctc	4449
Db	4381	AATTACTTAAAAACATTCACAGCAAAATTTGGAGATTTTAAAAAAAATATACATTTCCGTT	4440
Oy	4450	ggaattcagacactgtctaccogataataccaacaatctctctgaaccaagttctgtcc	4509
Db	4441	GGAAATTCGTGATTTATTCACAGATATATACCATATACTTAATGACAAAGTATCCCTTAGT	4500
Oy	4510	actgcagcgtgtgttcgaaaacctgcgcaaaacctgcgcagtcgacatctgcagcggcaac	4569
Db	4501	ACAGGATATGTTTTTGAAAAATCTTCCTAAACCGTTTTATCTAATTTACTTGATGGAAC	4560
Oy	4570	ctcgaaggcagctgtcgtacatctcccaagccaatgtcgtgaagaacagltgcgcccaagat	4629
Db	4561	TTGCAAGGTATATGTTTAAACATTTCCACACACCAATCGCTAAAAAACATGTGCCACAAAT	4620
Oy	4630	agcgaccttttcaggacatctgagcagcggcgaagaagtgcgaagttctctctgaactacaaa	4689
Db	4621	TCTGGATGTTTCAGCATTTTGATGATAAAGACAGATGTAAATGTTTATTAATATACAA	4680
Oy	4690	caagaaggaagataaagltgctgtgagaaacccaacccctaccttgcatagtaaaacaatltgcggg	4749
Db	4681	CAAGAAGGTGATTAATATGTTGAAAATTCAAATCCTACTTGTAACGAAATATATGTGTGA	4740
Oy	4750	tgtagccgcgatgtcctaattgcacgcgagaagaacagcggtcttaecggaaagaataacaca	4809
Db	4741	TGTGATGACAGATGCCAAATGTACCCAAAGAAATTCAGGTACCAACGGAAGAAATTCACA	4800
Oy	4810	tgcgagttactaaagcccgactccatccatcttcacgttcgacggagatttttgcaccgctc	4869
Db	4801	TGTGAATGTACTAAACCGATTCGATTTCTTATCTACTTTTCGATGTGATTTTCTGCAGTTCCCT	4860
Oy	4870	aatttccttgggcatctcctctcctgtgatacctcactgcatcgtcgtcatcgtlaccagcttactaca	4929
Db	4861	AACCTCTTAGGAATATCATCTTCTTATTAATACTCAATGTAATATATATACAGTTTCATTAA	4920
RESULT	6		
A04562			
LOCUS	A04562	5760 bp	DNA linear PAT 13-JUL-1993
DEFINITION	P.falciparum P.195 gene.		
ACCESSION	A04562		
VERSION	A04562.1	GI:410754	
KEYWORDS			
SOURCE			
ORGANISM	malaria parasite P. falciparum.		
REFERENCE	Plasmodium falciparum		
AUTHORS	Enayefot; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
	1 (bases 1 to 5760)		
	Holder,A.A., Sandhu,J.S., Odink,K.G., Lockyer,M.J. and		
	Riveros-Moreno, Y.		
TITLE	Cloning of DNA for protozoal antigens		
JOURNAL	Patent: BP 0154454-A-2 11-SEP-1985;		
FEATURES	THE WELLCOME FOUNDATION LIMITED		
SOURCE	location/Qualifiers		
	1..5760		

gene			/organism="Plasmodium falciparum"
CDS	216..5180	/db_xref="taxon:3833"	
	/gene="P_195"		
	216..5180		
	/gene="P_195"		
	/codon_start=1		
	/protein_id="CAA00367.1"		
	/translation="MKIIFFLCSPFFPIINTOCVTHESVOELVKKLEALDEAVLTGYS		
	LFQGEKRYLNENGTGATVATTSVPGSKGSVASGGSGVASGSAVSAAGSVAS		
	GGSGNSRWTNENSDSDSAKYADILKIRANVYLLTLEKLEPOLPLTNHMLTCDN		
	IHGKRYLIDGEEIENELLYKLNFPEYEDLLPRKINENVCANDQIPEFNIRANEDVILK		
	KLIVAGYRLPDINIKDNYGKMEDYIKKKKKTETENNELIESSKITDKNKNADEEEK		
	KLYAOVDLSYINKOLEBAHNLIVLEKRIDTLKKNNIKELDKINIRNPANG		
	NTPATLDRNKLIEHEKELEIKATIKFNPDLS..FTPLDELYELRKNKNIIDISAKV		
	EKTSKEPENEYNGVITYPLSYNDINMLNENSGDILINPDYERESKNIYNDISARV		
	KFIMIEKIKIEKIEKIESDKSYEDRSKSLDITTKYEKLINLIDYSKNNINDEKLN		
	PEKMAKRSYKVEKILKINPTFASYSKMSHLEKTLVAKLKMEDYSLRNIVKEKLN		
	YKNLSIKENIENIELTVENIKKDEDEQLPEFKTITQENKPDDEKILVESIVVAYQV		
	MNKILDELKTLQILKINYEIKNHINHPVSYOENQEPYVILVLEKIDILKIVMP		
	SLINEKKNIKTQCSQSDNSEPTSGELITQATYTPRQOAGSALEGDSVQAOQOROKA		
	QEPVPEVPEAKAQPVPPEAPVNNKTNVSLDLEKLYOLPLNLYTICHTILVSHST		
	MNEKILKQYITKEEBSKLESCDPLDLFPNQQNI PVM SFPDLSLYINHYMLVSK		
	EIGVGFILMLIEYKEMVCNLYKLKMDKILNLEEKAKYSIKVLTSSSMOPLSLTP		
	PODPEVSANDDTSHSNLNNSLILFNILNLSIGKNINYOGLIIOKSENEYEKILDK		
	SDPTNSETPTNFEVRSKADIDINSLNDESKRKLEIDIKKLTQDLSIDLVNKKYLET		
	RLPFRKRTYGVKQYKQIKIKTLIKPOLSKSLNPNKHYLONSPVFPNNKKKGAETAT		
	ENTLENTKILKHKIKLYKTYNGESSPLKTLSESIOTEDNYASLENFKYLSKIEGL		
	KDNINLEKRSKLSYLSRGLIHLIALEKVIKNNKNTGSPSNNTDNNALNLESTKFLP		
	EGDTAVATVSESGSDTLEQSQPKRPASTHGAESNTITTSQNVDDDEVDAIIVLIFEE		
	SEEDYEDDLGOVYTFEAAVTVSYDINILSKINERYELVILKPLAGYRSKLKOLENNVAT		
	FENVAVDILNLSRFRKRNFKNVLESDLIPYKDLTSSNVVAKDPYKFLNKEKRFLESS		
	YNYIKDSIDIDINPANDVIGYKYLISERSKSLDLSIKKTYINDGKEMKLPFLNIEE		
	TIKTYVDKIDILFVYHLKAVIKANTYKRSNVEYIKLELYNTKTODKLADPKNNRV		
	GIADLSDIYNHNNMLLKFSLSTGWFENULMSVLSNLDMLKARIYKVFPTPMKRTYI		
	QOOSGCFRHLDRECECKLNTYKQSGSKCVENSPNTENNNGGDDADAKCTEEDSGSN		
	GKKTKCECTKDCVPLSMVIFGSSNSNLGIFSLIMLILYSFI"		
BASE COUNT	2565 a 630 c 725 g 1840 t		
ORIGIN			
Query Match	54.8% Score 2706.2; DB 6; Length 5760;		
Best Local Similarity	72.1% P: 0; Mismatches 1343; Indels 45; Gaps 3;		
Matches 3587; Conservative			
Qy 10	atgaaatcaattcttcctcctcgttcattcttcgttttattatcaataaccacgaagtg	69	
Db 216	atgaagatcatattcttttttatgttcatctttttttttattatataaacacaaatg	275	
Qy 70	accacgaatccatcagagagctggttaagaacatgaaacttggaaagtcgcctc	129	
Db 276	acacatgaaagttatcaagaacttctcaaaaaaacatgaagactttgagaagatgc	335	
Qy 130	accgatacagctgttcacagaaggaagaagtgtgcgaatgaaggaaggaagtg	189	
Db 336	acaggtttatagtttttcaaaagaaaaaaatgatttaaatgaaggaacaaatg	395	
Qy 190	gccgttacaacacagacaccccggttcttaagagtgcttgagtcagcgtgcgc	249	
Db 396	gctgtttacacagatgacacactggttcaaaaggttcattggttcagatgcgtt	455	
Qy 250	tctgtgaccttgaggttctcgctgcctcggcggaagcgttgatcaagtgtgct	309	
Db 456	tcagttgcttcagcggcgtcagttgcttcacagtggttcagtggtgcacagtt	515	
Qy 310	gcaagcgcgcgttccgggaaacagtcgaagaacaaatcatctgacaaactcag	369	
Db 516	gcttcaggtgcttcagcgaatattcaagacagctacaaatctttcagatattca	575	
Qy 370	gacgcgaagctccacgcgaacctcaagcagcagtgagaaactatctccatac	429	
Db 576	catgctaaatcatctagcgtgatttaaaacacagacgaataattctgtaacat	635	

QY 430 gagctgaagtacccacagttgttcgcacctcaatacatactgctgcaactgtgtgaac 489
DB 636 GAACCTCAAAATATCCTCAACTCTTTGATTAACTAATCATATGTTAACTTGTGTGATAT 695
QY 490 attcattgcttcaatatcgtattgacggttgaagagatcaatgaacccgtacaag 549
DB 696 ATTCATGAGTTTCAAAATATTTAATGATGATGATGAAGAAATTAATGATTAATTAATAA 755
QY 550 ttgaattctactgactgctgcaaggccaactgaatgacttgcgcgaactgacat 609
DB 756 TTAACTTTATTTTGTATTATTAAAGCAAAATTTAATGATGATGCTAATGATAT 815
QY 610 tgcataattccattcaatttgaagatcagaagccaaggttggacglatitgaaggttg 669
DB 816 TGTCAATATGCTTTCAATCATCTTAAATTCGTGCAATGATTAAGCCTTAATAAACTT 875
QY 670 gtcttcggtatctgcgaagctctgcgacaatacaaggaacatgtggaaagatggaagat 729
DB 876 GTGTTGGATATGAAAGACCATTAAGCAATATTAAGATATGTAGCAAAATGGAAGAT 935
QY 730 tatattaaaaagaataagaagaccatcgagaacattgaagctgtgactcgagaatccaa 789
DB 936 TACATTTAAAAATTAATAAAACCATTAAGAAATATTAATGATTAATTTGAAGAAAGTAA 995
QY 790 aagacacatagacaaaataaagaatgcaacccaagaggaagaaagaagaagatgttaccag 849
DB 996 AAACATATGATTAATAATGAATGCACTAAAGAAAGAAAGAAAAATATATATACCA 1055
QY 850 gcccaatagcactgtccatctatacaaacgctgtgaagaagccataaccatcagc 909
DB 1056 GCTCAATATGATCTTCTATATTACAAATTAACAAATTAAGAACACATAATTTAATAAGC 1115
QY 910 gtactgagaagagcgtacagccctcaagaagaatgaataatcaagaagctgcagc 969
DB 1116 GTTTTGAAGAAAAACGTATGCACTTTTAAAAAAATCAAAACATTAAGGATTAATCTTGAT 1175
QY 970 aagattaatgaatlaagaatcctgcgcagccaactctggaagacacccctaaacgctg 1029
DB 1176 AAGATTAATGAAATTAATAATTCCTCCACCGCCCAATCTCGAATATACCAAAATACCTCTC 1235
QY 1030 ctgagacaagaagaagaatagaagagcagaagaagaagatcaagaagatcgcaaac 1089
DB 1236 CTTGATTAAGAAACAAAAATCGAGAACACGAAAAAATAAAGAAATTTGCCAAACT 1295
QY 1090 attaatctcaacatagattctcttactgactcccttgagctggaactactctaga 1149
DB 1296 ATTAAATTTAATATGTGATGATTTATTTACTGATCCACTTGAATTAAGATATCTATTAGA 1355
QY 1150 gagaagaataagaatatagacatctccgcgaagctcgagacaagaagaatcaacggaact 1209
DB 1356 GAAAAAAATTAATAATTTGATTAAGTCGCAAAAGTTGAAACAAAGCAATCAACTGAACCC 1415
QY 1210 aatgaatatcccaatggtgtgagcgtaccctctgtctataagatatcaacaagctctc 1269
DB 1416 AATGAATATTCCAATGATGATTAATCTTTGATTAAGCAATTAATTAACATGCTTTA 1475
QY 1270 aatgaatcacaatgcttcggtgacttgattgaacccctcgattataagaagaacccctc 1329
DB 1476 AATGAACCTTAATCTTTTGGTGATTAATTAATTCATTCATTTATATACAAAGAACAGT 1535
QY 1330 aagaatatctacacagacaatlgagagaagaagttatcaagaagaatcaagagaagatc 1389
DB 1536 AAAAATATATATGATGATTAAGAAAGAAATTCATTAATGAATTAAGAAAAAAT 1595
QY 1390 aaaaatlgagaagaagaatltgagagtgcagaagaagaagttacagaagccgcagaaagt 1449
DB 1596 AAAATAGAAAAAATAAATTTGAATCTGATTAATAAATCTTAAGAGCAATCTAACTCT 1655
QY 1450 ctacagatatcactaagaagtgaagaagctgctgaagagatctatgtctcaaatc 1509
DB 1656 TTAAATGATATTAACAAAGAAATATGAAAAATTTACTTATATGAATTTATGTATGACAAATTC 1715

QY 1510 aacaataacatgcgacctgacccaactcgagaanaatgtatggtgaagaacggttactctcaaa 1569
DB 1716 AATATATATATATGATTAATTAATTCGAAAAAATGATGAGTGAAGATTAATCATATTA 1775
QY 1570 gtgagaagaactgcacacataacattgacttgcctcctatgagaactctgaacatactt 1629
DB 1776 GTTGAGAAACTTACACACCCTTAATTAATTTGCAATCTGAGAAAAATTTCAACATTAATCT 1835
QY 1630 gagaagctcacaagaactcttaaglatatlgaggagactatctctgcggaacatgtgtg 1689
DB 1836 GAAAGTTTAACAAAGCTCTTAATATATATGAGAGATTAATCTTAAGAAATATATAGTAT 1895
QY 1690 gagaagaactaagaatgattacaagaatctcataagtaagaatgcgaagaagcgcgaag 1749
DB 1896 GAAAAAGAAATTAATAATTAATAAATTTAATATAGCAAAATTAATAAATGATTTGAAGA 1955
QY 1750 ctgtttgagacattgaagaagatgagaagacagttgttgagaagaagaattacaagaag 1809
DB 1956 TTAGTTGAAATATATTAATAAAGATGAGAGACACTTTTGAAAAAAATTTACTTAAGAC 2015
QY 1810 gaaataaaccagatgagaagatccttgaggtctcgataltgtltaagtcgaagtcag 1869
DB 2016 GAAATATTAACACAGATGAAAAATTTTGAAGTATCTGACATTTGTAAGATCAAGTTCA 2075
QY 1870 aaggtgctccctacatgacaagaatltgaactgaactaagaagaactcaactcgtgaagac 1929
DB 2076 AAAGTTTATTAATGAACAAATTTGACGAATTAATAAAGCTCAATGATTTTAAAAAT 2135
QY 1930 gtgaggtlaaacaataataatatactgtgcgaatgaattatbaagcaggagaataagcagaa 1989
DB 2136 GTAGAAATTAATAACATATATATACATGTCTCCAAATCTTACAAACAGAAATTAAGCA 2195
QY 1990 ccaataactctacatcgtactcaagaagaagatagacaagaactgaaagtgttcatgccc 2049
DB 2196 CTTATTAATTAATTTGTTGTTGAAAAAGAAATGATTAATTAATAAGTTCATGCTTAAG 2255
QY 2050 gtcgagaagcctgatacaagaagaagaagaacataataaactgaagagcgtcaagataac 2109
DB 2256 GTAGAAATCATTTGATTAATGATGAAGAAAAAATAAATAACACAAAGCTCAATGGATTAAT 2315
QY 2110 tccgagccttcacagaagaagagatataccgagcagctacacccaagcccggaacaag 2169
DB 2316 TCGGAACCATCAACCGAAGAGATAATTAACAGCAACAACTGACCAACCA 2375
QY 2170 gccggttcaagctctcgaagcgaatagcgtgcgaagctcaagcacaagaagcagaagca 2229
DB 2376 GCAAGATCTGCTTTAGAGAGATTCAGTACAGCAACAAACCAAGCAAAAAAAGCA 2435
QY 2230 cagctccagtgcccaagtgcgcgttccagaggtcgaagctcaagctcaagctcaacccaagct 2289
DB 2436 CAACCAACAGTACCAAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2495
QY 2290 cctgtgaataacaagaagcaggaatgtcaagaagaactgaactccttgagaagcctctag 2349
DB 2496 CCAATTAATTAATAACTGAAAAATGTTTCCAATTAATTAATTAATTAATTAATTAATTA 2555
QY 2350 tctctgaataacatccatcactctgcacaaaatatactcgtctcctcaacagcactatgaac 2409
DB 2556 TTTTAAATATCTTCATATATATATGTCACAAATATATTTTGGTTTACACCTCACTATGAAC 2615
QY 2410 gagaagaatcttaacaagatcaagaatacaagaagaagaagatgaactcgtcctctgt 2469
DB 2616 GAAAGATTAATTAATAAATTAATAAATTAATAAAGAGCAAAATTAATTAATTAATTAAT 2675
QY 2470 gatccactgagcctgctgttcaatccatccagaagaacatccgcttgaatctatgctc 2529
DB 2676 GATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2735
QY 2530 gataagctcaaca-----atct 2547
DB 2736 GATAGCTTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2795
QY 2548 ctctctcaactgtlcatlgagatatatgagaagaagatggtctgcgaacctgtataaactc 2607

D 2796 GTATTATATATCTTATGGAATTTATGAAATAAGAAATGTTCTTAATTTATATAACTT 2855
Q 2608 aaagacaagacaagaaltaagaacctctcggaagagctaagaagctccacctgtt 2667
D 2856 AAGATATATGCAAAATTTATTTATAGAGAGCGAAAAAGATATCCATCTGTGA 2915
Q 2668 aaactctctccagctccatccgaacctgtctccacacctgaagacaagccgaa 2727
D 2916 AAACTCTTTTCAAGTTATCATCATCAACCATTTATCATTAACCTCAGATTAACCCGAA 2975
Q 2728 gtgagcgctaaagcagacacctctccacctgaacctlaaactaactgaactgtt 2787
D 2976 GTAGTGCAATATGATGATACATCATCTTCAAAATTTGAATATATAGTTAAATATTT 3035
Q 2788 gagaacatctgtctctcgcaagaataagaacatctaccagaactatctgacagaa 2847
D 3036 GAAAACATATTTGAGCTTTGGAAAAACAAAATATATACCAAGAAATTAATGCTCAAAA 3095
Q 2848 tcgtccgagacctctcagagaagatactgaagacagcgacacatctctaagcagagc 2907
D 3096 AGTAGTAAACCTTTTATGAAAGATATTTAAGATAGTGATATTTATATGAATCT 3155
Q 2908 ttcaactactcgtaaatctaaagccgatgatalcaactctcttaacgatlaaactaa 2967
D 3156 TTTCAATATTTTGTAAATCTAAAGCTGATGATATTAATCATTTGAATGATGATCAAA 3215
Q 2968 cgtagaagagcttgagaagggacataaataagctgaagaagacatcgacgtcgac 3027
D 3216 AGGAATCAATTTAGAAAGATATTAATAATTAATAAAACCTTTACAGTTATCTTGAT 3275
Q 3028 ctgtaacaagaactaacaactgaactgaagagacctctcgacaagaagaagacgtcgac 3087
D 3276 TTATATATATATATTAATTTAAATTTAGAAAGATTTTGTATTAATAAAGAAACAGTGT 3335
Q 3088 aagataagatcgagatcaagaagctgactgtctcgaagagcgactgaagaactc 3147
D 3336 AAAATATTAATGCAAAATTAATAAAACCTTATTAATAAGAAATTAAGAAATCAAAATG 3395
Q 3148 aacctgaagacaatccgaagacgtaactcgagagacctctgaagctcttaagaagaag 3207
D 3396 AATTCACCTTAATTAACCAAGACATGATTAACAACCTTTCTGTCTTTTACCAAAAAA 3455
Q 3208 aaggaagccgagatcgcgcgagacagagaacacatctlgagaacacacaagatctctca 3267
D 3456 AAAGAAGCTGAATATAGCAGAAACGAAACACATTAGAAACACAAAAATTTATTGAA 3515
Q 3268 caactaagaagcctcgtaagtalataatlgcgagctctctcctgaagactctcc 3327
D 3516 CATTAATTAAGACATTTGTTAAATTAATTAATGTAATCATCTCCATTAATAAACTTTAAGT 3575
Q 3328 gaagagagcatccgacgcgagagataactagccagcctcgagaactccaagctctgt 3387
D 3576 GAAATCAATCAATTAACAGAGAAATTAATTAATTCAGTTTAAAGTTTAAAGT 3635
Q 3388 aagctcgagaagcagctggaagacaactggaactlgagaagaagaagctcagctacctc 3447
D 3636 AAATTAAGAGAAATTAAGATTAATTAATTAAGAAAAAGAAAAATTTATCTACTTA 3695
Q 3448 tctgagagactgcatcaactgtaactcgccgagctcaaggaagctcaltlaagacaagaact 3507
D 3696 TCAAGAGAGTTTACATCTTTAATTTGCTGAATTAAGAGAAATTAATAAAATTAATAT 3755
Q 3508 accggaatagcccaagcaggaataataacagagctgaaataacgcatcggaactctacaag 3567
D 3756 ACAGAGATTTCTCCACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3815
Q 3568 aagctcctcgctgaagacaagatgctgcacatgctgtgtctgaactgagctccgacaca 3627
D 3816 AAATTTTCTCCAGAGAACAGATGTTGCAACAGTTGTAAGTGAAGTGATCCGACACA 3875
Q 3628 ctggaagagctcaactaagaagcctgacatctcatgtcgagcgagatccaataca 3687
D 3876 TTAGAACAAAGTCACCAAGAAACACAGCATCACTCATGTAGAGCAGAGCTTAACACA 3935
Q 3688 attaccacatctcgaagcgtcgagagatgagtgatgacgtcatctgtgcatcttc 3747
D 3936 ATAAACAATCACAATAATGTCAGATGATGAATGATATACATTAATGATATATTT 3995
Q 3748 ggcgagagcgagagagctcgatgacatcgccgaagtggccaagctgaagctgtgact 3807
D 3996 GGACAATCCAGAGAGATTAATGATGATTTAGACAACTAGTAACAGAGAAAGCATTA 4055
Q 3808 ccttcgctgatalaacaatctgtccaataatcgagaacgaatacgaagtgtctatctg 3867
D 4056 ACTTCGTAATTTATTAATCATCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 4115
Q 3868 aaactctgcaagcgtctataagctctcgaagaagaagctlgagaataacgtagtacc 3927
D 4116 AAACCTTTAGAGGTGTTTATAGAAATTTAAAAAACAATTAAGAAATTAACGTTATGACA 4175
Q 3928 tcaatgtaacgctggaagacatctgaacagccgcttaataagagaagaatltcaag 3987
D 4176 TTTATGTTAATGTTAAGATATTTTAATTTACAGATTTAATAACGTGAATAATTCAA 4235
Q 3988 aagctctgagagcgactgtgattccctalaagaacctgaacctcttaactagctgtgc 4047
D 4236 AATGTTTAGATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4295
Q 4048 aaggaaccatacaagcttccatlaagaagaagggataattctgtctagtataac 4107
D 4296 AAAGATCCATTAATTAATTTCTTAATTAAGAAAAAGAAATTAATTTTAAGCAGTTAAT 4355
Q 4108 tatacaagacatccatcgacaacccatcaatctcgtatctgactgtgctgaggtatca 4167
D 4356 TATATTAAGATTAATTAATTAATTAATTAATTAATTTTCAAAATGATGTTCTGATATAT 4415
Q 4168 aagatcctgagcgaataatacaagctcgacctgactctatlaaagaatatacaagat 4227
D 4416 AAAATATTAATCCGAAAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4475
Q 4228 aagcaagcgagaatgaataatctgcacctctcgtatatacaatcgaaacctgtacaag 4287
D 4476 AAACAGAGTGAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4535
Q 4288 aagctgaagacaatccgacctctcgtatatacaatctgagcgaagctcactat 4347
D 4536 ACAGTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4595
Q 4348 acttaagagaagcaatgtagaagtaataatcaagaagctggaactcctccaacaact 4407
D 4596 ACATATGAGAAATCAACGTAAGATTAAATTAATAAACAATTAATTAATTAATTAATTA 4555
Q 4408 caagacaagctgagagattcaagaataatacaaatctgctgaatctgaagactgtc 4467
D 4556 CAAGACAAATTTGCGAGATTTTAAAAAATAATTAATTAATTAATTAATTAATTAATTA 4715
Q 4468 accgatataacccaacaactctcctgaaccaagcttctgccaactgagctgtgtcgaa 4527
D 4716 ACAGATTAATAACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4775
Q 4528 aactcgccaagaagctgtgagcaatctgtcga -cggaacctgcaagggcatgtgaa 4586
D 4776 AACTTGTCAATTAATCCGTTTATCTTAATTAATTAATTAATTAATTAATTAATTAATTA 4835
Q 4587 catctccgaagcacaatgctgtgaagaagctgtg -ccccagaatlaagcgctgttagg 4644
D 4836 CATTTCAACACCAATGCGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4895
Q 4645 catctggaagcagcggaagctgtgaagctgtcctgtaactcacaacaagaagaagataag 4704
D 4896 CATTTAGATGAAGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4955
Q 4705 tgcgtgagaagcccaacctactcgaatgaagaacaatggcggtgtgtagcgcatgtct 4764
D 4956 TGTGTTGAAAAATTCAAATCTACTTGTAAAGAAATTAATGATGATGATGATGATGATGATG 5015

QY	4765	aaatgcacccggaagagacagcgycgtcttaacggaagaataatcatatgcagtgylactaa	4824
DB	5016	AAATGTAAATCGAAGAGATTCAGGTAGCAACGGAAGAAATACATGTGAATGTATAA	5075
QY	4825	ccgcagctctatccactctctgacggagatttttctcagctcaatttcctgggcacc	4884
DB	5076	CTGATATTGTTATTCACACTTTTCATGATGTTAAATTTCTGACAGTCTCTTAACCTCTTAGGATAA	5135
QY	4885	tcctcttcgctgcatctcatctgcatctctgtaacagcttcatatctaataagatg	4939
DB	5136	TCATTCCTATTAATATCTCATGTATATATATACAGTTCATTTAAATAAATGTAGG	5190
RESULT	7		
LOCUS	E00656	5760 bp	RNA
DEFINITION	CDNA encoding P.195 protein of plasmodium falciparum.	linear	PAT 29-SEP-1997
ACCESSION	E00656		
VERSION	E00656.1	GI:2168935	
KEYWORDS	JP 1986019490-A/1.		
SOURCE	Plasmodium falciparum.		
ORGANISM	Plasmodium falciparum.		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5760) Ansonii,A.H., Malkuru,J.R., Jiyasublaa,S.S., Barenteina,R.M. and Kereu,G.O.		
TITLE	DNA CLONING FOR ANTIGEN OF PROTOZOA		
JOURNAL	Patent: JP 1986019490-A 1 28-JAN-1986;		
COMMENT	WELLCOME FOUND LTD:THE OS Plasmodium falciparum PN JP 1986019490-A/1 PD 28-JAN-1986 PE 22-FEB-1985 JP 1985034280 PR 22-FEB-1984 GB 84 8404992, 26-SEP-1984 GB 84 8424340 PI ANSONII AASAA HORODAA, MATKURU JIEMESU ROTSUKIYAA, PI JIYASUBIDIA SHINGU SANDEYU, BARENTEINA RIBEROSU MORENO, PI KERITSUTO ODEINKU PC C12N15/00,A61K39/015,C07H21/04,C07K13/00,C12N1/20,C12P21/02, PC (C12N1/20), PC C12R1:19),(C12P21/02,C12R1:19); CC strandedness: Single; CC topology: linear; CC hypothetical: No; CC anti-sense: No; FH key Location/Qualifiers FH CDS 216..5180 FT /product-'P.195 protein' FT sig_peptide 219..272 FT mat_peptide 273..5177 FT /product-'P.195 protein'. FEATURES source Location/Qualifiers 1..5760 /organism='Plasmodium falciparum' /db_xref='taxon:5833' BASE COUNT 2565 a 630 c 725 g 1840 t ORIGIN		
	Query Match	54.8%;	Score 2706.2; DB 6; Length 5760;
	Best Local Similarity	72.1%;	Pred. No. 0;
	Matches 3587; Conservative	0;	Mismatches 1343; Indels 45; Gaps 3;
QY	10	atgaataatcatttctctctctgcttcattcgtttttatataataactaagcgcg	69
DB	216	ATGAAGATCAATATTCCTTTTATGCTCATTCCTTTTTTATATATAACAAATGTGA	275
QY	70	acccagaaatcatcagcagcgtgttgaagaactggaagcttggaaatgcgcctt	129
DB	276	ACACATGAAAGATTATCAAGAACTGTGCACAAAACCTAGAGCGTTTGAAGATGCAGTATTG	335
QY	130	accgataacagcgttccagaagaagaagatgctgctgaatgaaaggagcagtgacag	189

Dd	336	ACAGGTTATAGCTTTATTTTCAAAAGGAAAAATGGTATTTAAATATGAAGAAACAAGTGAACA	395
Qy	190	gccggttaacacgacacaccccggtctcaaaaggctctgtgtctaaagctgcggtctcggttgg	249
Dd	396	GCTGTTACAACTAGTACACCCGTGGTTCAAAAGGTTCAAGTTCCTTCAGTGGTTCCAGGTGC	455
Qy	250	tctgtgaccttgggggttccgtgcctccgacggcagctggcatcaagtgtgctcaagt	309
Dd	456	TCAGTTGCTTCAGGTGGCTCAGTGGCTTCACAGTGGCTCACTGTCCTTCAGGTGGCTCAGTT	515
Qy	310	gcaagcggcgggttccgggaacagtcggaagacccaatccatctgacaaacctctagccatccc	369
Dd	516	GCTTCAGGTGGTTAGGTAAATTCAAGACGTCACAAATCTTCAGATTAATTTCAAGTGATTCA	575
Qy	370	gaagcgaagctccacgcgcgaactcaagcacccagttgaaagaaatactctccactcaacaag	429
Dd	576	GATGCTAAATCTTACGGTATTTAAAACACAGAGTACGAATATTAATCTTTAACTATCA	635
Qy	430	gaagctgaagatcccaacagttgttcgacctcaactacatctatgtctgacatggtgataac	489
Dd	636	GAACGCAAAATTCCTCAACCTTGGATTTAACATCAATATGTAACTTTGGTGATTAAT	695
Qy	490	attcattgctccaatatctgatctgaagctgtacgaagagatcaatgaactccctgtacaag	549
Dd	696	ATTCAATGGTTTCAAAATTTTAATGTAGATGATGAGAAATTAATGAATTAATATATAAA	755
Qy	550	ttgaattctactctgcagctgtgtaaggccacaactgaatgaagctgttcggccaatgactat	609
Dd	756	TTAAACCTTTATTTTGGATTATTTAAGAGCAAAATTTAAATGATATGTATGCTATATGATTA	815
Qy	610	tgtaaatctcatccaattcttgaagatcaagacacagagttgagcttatctgaagaagtgtg	669
Dd	816	TGTCAAAATACCTTTCAATCTTAAATTCGTGCAATGSAATTAGACCTACTTAAAAACTT	875
Qy	670	gtcttcggaatctgcgaagcctctgcgaacatcaagagcaatgttgggaagaatgtgaagat	729
Dd	876	GTGTTGGATTTAAGAAAAACCTTAGCAATATTTAAAGATATGTAGAAAAATGCAAGAT	935
Qy	730	tatatataaagaatlaagaaagccatctgcgaacatctaaagagctgtctgaagaatcccaaa	789
Dd	936	TACATTTAAAAAAATTTAAAAAACCATTTGAAAAATTTAAATGAATTAATTAAGAAAGTGAAG	995
Qy	790	aagaccatagacaaaataaagaatgcgaacacgaagaggaagaagaagaagagttgtacag	849
Dd	996	AAAACAATTTGATTAATAATTAAGCAATGCACTTAAAGCAAGCAAAAAAATTTATRCCA	1055
Qy	850	gcccgatcgagcctgtccatctataacaacaagcttgaagaagccataacctcatcagc	909
Dd	1056	GCTCAATTTGATCTTTCTTATTTTAAATTAACAAATTTGAAGAGACACATTAATTTAATAGC	1115
Qy	910	gtacgtgggaagcgcatagacacccctaaagaagaatgaagaatctgaagaactgtctgac	969
Dd	1116	GTTTTAGAAAAACGTAATTTGACCTTTTAAAAAAAATGAAGAACTTTAGGAATTAATCTTGAT	1175
Qy	970	aagatttaatgaatlaagaaatcctctcgacacgaacactcttggagacacccctlaacacgtgt	1025
Dd	1176	AAGATTAATTTGAATTTAAAAATCCCCACCGCCAAATCTTGGAATATCACCAATTAATCTTC	1235
Qy	1030	ctgagcaagaacaaagaatagagaggaacgagaagaagagatcaagaagatctgcgcaaac	1089
Dd	1236	CTTGATTAAGAACAAAAAATTCGAGGAACAGAAAAAGAAATTAAGAAATTTCCCAAACT	1295
Qy	1090	attaagtccaatagatctcctctttacagatccccccttgaagcttgaagatcgaacttggga	1145
Dd	1296	ATTAAATTTAATTTAGTTAGTTTATTTACTGATCCACTTGAATTTAGAAATTAATTTAAGA	1355
Qy	1150	gagaagaataaagatatagacatctccgcgaagctcgagacaaaggaatcaacgcgaacct	1205
Dd	1356	GAAAAAAATTAATAATTTTGAATTTATAGTCGAAAGGTTGAAACAAAGGAATCAACTGACCC	1415
Qy	1210	aatgaatatcccaatggtgtgaagcttacctgtcttataacgatatcaacaagcgtctc	1265

Dh 1416 AATGAATATCCAAATGAGTACTTATCTTGTGCATATTAACGATATTAAACATGCTTTA 1475
Qy 1270 aagcgaccatagcttcggtgacttgatcacccttcgattacgaagaaccctc 1329
Dh 1476 AATGAACCTTAATCTTTGGTGATTTAATTAATCCATTGTGATTAACAAAAGAACCAAGT 1535
Qy 1330 aagaatctcacagacaagagagaagaagttatcaacgaatcaagaagaagtc 1389
Dh 1536 AAAAACAATATACTGATTAATGAAGAAAAAATTCATTAATGAAATTAAGAAAAAATTT 1595
Qy 1390 aaattcgagaagaagaattgagatgacaagaagaattcagaaccgcgaagaagt 1449
Dh 1596 AAAATAGAAAAAATAAATGAATCTGATAAAAAATCTTGAGAAACAGATCTTAAGTCT 1655
Qy 1450 ctaacgatcactaagaagatlgaaagctgctgaaacgagatcagatccaaatc 1509
Dh 1656 TTTAAATGATTAACAAAAGATATGAAAAATTAATTAATGAAATTTATGATAGCAAAATTC 1715
Qy 1510 aacataacatcgaccctcgaccactcgagaagaatlgatgggaaaaacgctactctaca 1569
Dh 1716 AATATATATATAGATTAACTTAATTTGAAAAAATGATGGGTAAAAAGATATTCATATAA 1775
Qy 1570 gtggaagaactcgacacacataaactctgcatcctatgagaattcctaagataactc 1629
Dh 1776 GTTGAGAAACTTACACACCTTAATCTTTCATCTCTATGAAAAATCTTAACATATCTT 1835
Qy 1630 gagagcctcaccaagccttaagtatactgagactatctcgcggaacatgctgtg 1689
Dh 1836 GAAAAAGTTAAACAAAGCTTTAAATATATGAAAGATTAATCTTTAAAGAAATATAGTACTT 1895
Qy 1690 gagaagaactcaagattatcaagaagatcctaaglaagtcgacaagaacgagacg 1749
Dh 1896 GAAAAAGATTTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1955
Qy 1750 ctgttgaagaacatlaagaagatgagaacagctgttctggaagaagatcaacaagac 1809
Dh 1956 TTACTTTAAATATTTAAAAAAGATGAAACAGCTTTTGGAAAAAATAATTAATTAATTA 2015
Qy 1810 gaaataaaccagatlgagaagaatctcgtgagctccgcatgtttaagttcaagttgag 1869
Dh 2016 GAAATATTAACCGAATGAAAAAATTTTAAAGATATCTGCATTTGTAAGTAATGCAAGTTCAA 2075
Qy 1870 aaggtgctccatcgaaagaatlgatgaactcaagaagaactcaactctcgaagaac 1929
Dh 2076 AAAGTTTATTAATGAACAAATTTGACGATTTAAAAAAGACTCAATTAATTTTAAAAAAT 2135
Qy 1930 gtggaagttaaaacataatatacatgctgcgaatagttataagcagagataaagaa 1989
Dh 2136 GTAGAATTAATAACATTAATATCATGTCTCCAAATCTTTACAAACAAAGAAATTAAGCAAGAA 2195
Qy 1990 ccatactaccatcgctactcaagaagaagatagacaacacgaaagtctcatgccc 2049
Dh 2196 CCTTAATTAATTAATTTGCTTGAAGAAAGAAATTTGTAATTAATTAATTAATTAATTA 2255
Qy 2050 gtctgagagccgtgacaaagagagaagaagacatlaaactgagagcagctcagataac 2109
Dh 2256 GTAGAATATCTGATTAATGAAGAAAAAATAATTAATTAATTAATTAATTAATTAATTA 2315
Qy 2110 tccgaagcctccagagaagaagataaaccggaagcgtacacccaagcccggaacaag 2169
Dh 2316 TCGGAACCATTAACCGAAGGAGAAATTAACAGCAACCACTTACAAAACCTGCACAAACAA 2375
Qy 2170 ggcggttcagcttcgaagagatagctgcaagctcaagacagaagcagaagagca 2229
Dh 2376 GCAGAGATCTGCTTTAGAGGAGATTCACTACAGCAACAGCAACAGAAACAAAACAAACCA 2435
Qy 2230 cagcctcagatgacagctgcggtcccaagagctaaagctcaagttgctcaccacagct 2289
Dh 2436 CAACCAACAGTACCGATCCAGTACCAACAAACAAAGCAACAGTCCCAACACACACACCA 2495
Qy 2290 cctgtgaataaagaaccggaagtgtcagcaaaactggaactccttgagaagcctctatgag 2349
Dh 2496 CCAGTAATATTAATAAATGAAAAATGTTCCAAATTAATTAATTAATTAATTAATTAATTA 2555

Qy 2350 ttccgtgaatcacatctacatctgcacaaatatccctcgctctcacaagcactatgaac 2409
Dh 2556 TTTTAAATTAATCTCATATATATGTCACAAATTAATTTTGGTTTACACTCAACTATATGAC 2615
Qy 2410 gagaagatcttcaaacagtlacaagaatacacaagaagaagagaglaaactgctcctgt 2469
Dh 2616 GAAAAATATTTAAACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2675
Qy 2470 gatccacgtgacccgtgcttcaatataccagaacaacatcccgctatgatlctcatgttc 2529
Dh 2676 GATCCATTAAGCTTAATGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2735
Qy 2530 gatagccctcaaca-----atctc 2547
Dh 2736 GATAGCTTTAAACATATGTTTATACACACTATTTATGGCTTTATGAAAAAATAATGGTTAT 2795
Qy 2548 ctctctcaactgctcatalgagatataatgagaagagatgctgcgaactglataaactc 2607
Dh 2796 GTATTTATATTAATTAATGAAATTTATGAAAAAGAAATGGTTGTAATTTATTAATTAATTT 2855
Qy 2608 aaagaacaagcaagaatlaagaacctctcgagaagaactaagaagctccactcgtt 2667
Dh 2856 AAGGATTAATGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2915
Qy 2668 aaactctctctccagctccatccacaccactgctctcacactcaagaagaagccgga 2727
Dh 2916 AAACCTCTTTAAAGTTTCATCAATGCAACCATTAATTAATTAATTAATTAATTAATTAATTA 2975
Qy 2728 gtgagcgttcaacgacgaacctctcaactcgacacactlaaactcactgaaactgctt 2787
Dh 2976 GTAAGTCGAAATGATGATACATCATCTCAATTTGAATTAATTAATTAATTAATTAATTT 3035
Qy 2788 gagaacatccgtctctcgtcggaagaatlaagaacatctccagaagactatgtgacagaa 2847
Dh 3036 GAAAAATATTTAGCTTTGGGAAAAAACAATAATTAATTAATTAATTAATTAATTAATTAATTA 3095
Qy 2848 tgcgtcgagaactctcaagagaatactgaaagaacgcgaacatctcaacgaagagc 2907
Dh 3096 AGTAGTGAACCTTTTATGAAGAAGTATTAATAAGTATGATACATTTTATTAATGAATCT 3155
Qy 2908 ttcaactaactcgtgtaaaactcaaaagccgaatgatalcaactctctcaacgtatgaactcaa 2967
Dh 3156 TTTACAAATTTTGTAAATTTCTAAAGCTGATGATTAATTAATTAATTAATTAATTAATTAATTA 3215
Qy 2968 cgtlaagaagctcggaagagacataaataagctggaagaagacacgtgaactgagac 3027
Dh 3216 AGGAAGAAATTTAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3275
Qy 3028 ctgtacaacaagtlacaactgaaactggaagagactctcgaacaaagaagaagcagctgcgc 3087
Dh 3276 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3335
Qy 3088 aagtlataagatgcagatcaagaagtlgaactcgtcgaagagcagctlgaaagcaactc 3147
Dh 3336 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3395
Qy 3148 aactcactgaaacatccgaacacgltactcgagaactctcagtgcttccatacaagaag 3207
Dh 3396 AATTCACCTTAATTAACCCCAACCATGTATTAACAAACCTTTCTGTTTCTTTAAACAAAAA 3455
Qy 3208 aaggaagccgagatcgccgagaacagaagaacactctggaagacacacaagactctctcaa 3267
Dh 3456 AAAGAGCTGAATTAATGCGAAATGGAANACCATTAAGAAACCAAAAAATTTATTGANA 3515
Qy 3268 cactacaagaagcctcgltcaagtatataatgctgagagcttctcctcgaagactcctcc 3327
Dh 3516 CATTAATAAGACCTTTGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3575
Qy 3328 gagagagacatccagaaccgagataactacgcagcctcgagaactcacaaggtcctgtct 3387
Dh 3576 GAAGATCAATTAATTAACAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3635

OY	3388	aaactcgaaggcgaagcgcgaagcgaacccggaacctcggaaagaagaagcctgaagctc	3447
Db	3636	AAATTAGAGAGAAAATTAAAGCATTAATCTTAATTTTGAAAAGAAAAAATTTATCATACTTA	3655
OY	3448	ctcagcggaactgcaacactgcatcgccgagctcaaggaagtcataagaacaagaactac	3507
Db	3696	TCAAGAGGTTTACATCTTTTAATTTGGTGAATTTAAAAGAAAGTAATTAATAAATAATTAAT	3755
OY	3508	aacggcaatagcccaagcggagaataatacagaagtgaaataagacgtgaatccttaag	3567
Db	3736	ACAGGTAAATTCTCCAACCGTAAATTAATACGGATGTTTAACAAATGCATTAGAACTTTACAA	3815
OY	3568	aagtcctcgtcgaagaaacaagatgctgcgaactggtgtctgaatctcgtcgcagaca	3627
Db	3816	AAATTTCTCCAGAGAGAACAGATGTTGCACAGTTGTAAAGAAAGGATGCCAGACA	3875
OY	3628	cttgagagagctcaactaagaagccttgatctactacatgctggagccgagttcaataca	3687
Db	3876	TTAGACAAAGTCAACCAAGAAAGAACGACATCAACTCATGATGAGAGCGAGAGCTTAACCA	3935
OY	3688	attacaacatctcgaacgcgcgaactgagatgagtcgaatgcatactatgtgcctactc	3747
Db	3936	ATTAACAATCATCAAAATGTGCATGATGAAGATGATGATGAGCTAATCATGTACTATATT	3995
OY	3748	ggcgaagcgaaggagacctcgaatgacctcgcgcgaagtgtcaacgggtgaagctgtcaat	3807
Db	3996	GGAGATATCCGAAGAGATTTGTGATGATTTAGCAAGTGTGAACGGAGAACCACTAATCT	4055
OY	3808	ccttcgcgatgtgaatacattctgctccaatctgagaacgaatacgaagtctcactgt	3867
Db	4056	ACTTCGGTAATTGATTAACATACCTTTCTAAATTTGAAAATGAAATGATGAGGTTTAAATTA	4115
OY	3868	aaacctcgtcgaaggcgtcatagttctcgaagaaacagctggagaataacgctgac	3927
Db	4116	AAACCTTTAGAGGCTGTTTATAGAGTTTAAAAAAACAATTTGAAAATAACCTTATGACA	4175
OY	3928	ttcatgtcaacgcggaaggaactctcgaacgcgcgtcttaataagagaagaaattcaag	3987
Db	4176	TTTATGTGTAATGTTAAAGCATTTTAAATTTAAATTCACGATTTAAATTAACGTGMAATTTCAA	4235
OY	3988	aacgctctggaagcgaactgattccctcaaaagacctgaacctcctctaactaagttgc	4047
Db	4236	AATGTTTAGATACATGATTTAAATTTCCATATAAAGATTTAACTCAAGTAATATATGTTGTC	4295
OY	4048	aaggaacctatacaagttccctcaataagaagaagagatgaataattctcgtctaagtacac	4107
Db	4296	AAAGTTCATATTAATTTCTTAATTAAGAAAAAGCATTAATTTCTTAAGCAGTTATATAT	4355
OY	4108	tataccaagactccatcgcgaacccgatataactttcgtcaatgaatgtcgtgggtatcac	4167
Db	4356	TATATTAGGATTCAAATAGATACGGATTAATTAATTTGCAATGATGTTCTTGATATATAT	4415
OY	4168	aagatcctgagcgaataatacaagctctgaccttgacctatataaagtatatcaagat	4227
Db	4416	AAAATATTATCCGAAAAATTAATACAGATTTAGATTCAATTAATAAATAATATATCAACAC	4475
OY	4228	aagcaagcgcgaatgaataatactcgccttcctgaataaactcgaacccctgacag	4287
Db	4476	AAACAAGTGAATAATGGAATTAACCTTCCTTTTAAACAATATTGAGACCTTAATATATA	4535
OY	4288	acagatgaacgaacaaatcgacctcttcgtaatcacctcggagccaaggtcctcaactat	4347
Db	4536	ACAGTTATATGATTAATAATTTGATTTATTTGTAATCATTTTGAAGCAAAAGTTCTAATATAT	4595
OY	4348	acttaagagaagacaaatctggaagttaaatcaagagacgtgaactactcaaaacaatc	4407
Db	4596	ACATATGGAATATCAAAACGTATGAAGATTAAATTAAGAACTTAATTAATCTTAATAAACATTT	4655
OY	4408	caagcaagcgtgcgaattccaagaataatacaaatcttcgtcggaatgtcgaacgctgt	4467
Db	4656	CAAGCAAAATTGGCAGATTTTAAAAAAAATAACAAATTTGCTTGGAAATTGCTGATTTATACA	4715
OY	4468	accgatcttaaccacaacaaactctctcgaccgaattctcgtccaactgacagtgtgtcgaa	4527

Db	4716	ACAGATTATTAACCCATAAATACTTATGTGCAAGTTCCTTACTACGAGTATGTTTTTGA	4775
Qy	4528	aacctgcgcaaaaacagtcgtcgaagaatctgtcga-cggcaacctgcagggcagtcgta	4586
Db	4776	AACGTGCTAAAAATCCGTTTTTATCTAATTTACTGTGATTGGAACCTTGCAAGTATGTTTAA	4835
Qy	4587	catctccagagcaaccaatgcgtggaagaaacagtg--cccccaagataagcgctglttcagg	4644
Db	4836	CATTTCACAAACCCCAATGCTGTAAAAAAAACAATGATCCACAGAGTTCTGGAATGTTTTCGA	4895
Qy	4645	catctgcgcgagcgcggaagatgcgaatgctcgttcctgcgaactcaaacagaagaagataag	4704
Db	4896	CATTTAGTTGAAGAGGAAGATGTAAATGTTTATTAAATTACAAAACAAGAGTGTAA	4955
Qy	4705	tgctgtggaagacccaacacctacatctgcgaatgaacaacatggcgagtgtagcgcatgct	4764
Db	4956	TGTGTGAAAAATTCAAAATCCCTACTCTTGTACGAAAAATATATGGTGGATGTGATGCAGATCC	5015
Qy	4765	aaatgcacgcggaagagacagcgtcctaacggaagaaatacacatgcgagtgactaag	4824
Db	5016	AAATGTACCGAAGAAAGATTCAAGTGTAGCAACGAAAGAAATACATGTGAAATGTCTAA	5075
Qy	4825	ccggaactccatccaactcttcgaacggagatcttttgctccagctcatcaatcttcctgggaac	4884
Db	5076	CGTATTTGTTATCCACTTTCGATGGTAAATTTTTCGACATTCCTCTCAATCTCTTGGAGATA	5135
Qy	4885	tcccttcgcgcgcatccatcgtcatgctgatacctgtcaacagctcatcatatagatcgaag	4939
Db	5136	TCATCTCTTAATAAATACATCATGTATATATATATACAGTTCAATTTAAAAAATGTAGC	5190

RESULT	8	
PEP190		
LOCUS	5282 bp	DNA linear INV 21-FEB-1996
DEFINITION	P.falciparum gp190 (MSA1, MSP1, PMMSA) gene for merozoite surface antigens precursor.	
ACCESSION	X03371	
VERSION	X03371.1	GI:929797
KEYWORDS	gp190 gene; merozoite surface antigens; MSA1 gene; MSP1 gene; PMMSA gene.	
SOURCE	malaria parasite P. falciparum.	
ORGANISM	Plasmodium falciparum	
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS	1 (bases 1 to 5282)	
	Mckay,M., Goman,M., Bone,N., Hyde,J.E., Scalfe,J., Certa,U.,	
	Stuenkelberg,H. and Bujard,H.	
TITLE	Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level	
	EMBO J. 4 (13B), 3823-3829 (1985)	
JOURNAL	86136024	
MEDLINE	2 (bases 1 to 5282)	
REFERENCE	Pan,W., Tolle,R. and Bujard,H.	
AUTHORS	A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSA1	
TITLE	Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)	
JOURNAL	96123395	
MEDLINE	3 (bases 1 to 5282)	
REFERENCE	Bujard,H.	
AUTHORS	Direct Submission	
TITLE	Submitted (23-JUN-1995) H. Bujard, Zentrum fuer Molekulare Biologie der, Universitaet (ZMBH), Im Neuenheimer Feld 282, Heidelberg 69120, FRG	
JOURNAL	On Aug 5, 1995 this sequence version replaced gi:9920.	
COMMENT	Location/Qualifiers	
FEATURES	1..5282	
	/organism="Plasmodium falciparum"	
	/isolate="K1"	
	/db_xref="taxon:5833"	
	/chromosome="9"	
	/dev_stage="blood stage"	
	72..128	
	/gene="gp190, MSA1, MSP1, PMMSA"	
	sig_peptide	

```

gene      /note="putative"
           72..4964
           /gene="gpl90, MSAL, MSP1, PMMSA"
CDS       72..4964
           /gene="gpl90, MSAL, MSP1, PMMSA"
           /codon_start=1
           /product="precursor for the major merozoite surface
           antigens"
           /protein_id="CAA27070.1"
           /db_xref="GI:929798"
           /db_xref="SWISS-PROT:P04932"
           /translation="MKIIIFLCSFLFPIINTOCVTHESYOELVYKLEALBDVAVLTGYS
           LFHKEMILNEEETITTKGASAGSSTGSGTSGSGSGSPSSRSTMLPDRNSTSA
           SPSPADSDSAKSYADLKHVRNNTLTIKELKYPQLDLNNHMLTLDNTHGFEYLD
           GPEINELLYKINYPDLIRAKINDVANCOCYRPLKIRANLDVLYKLYGRYP
           LDNIRKDVGMKEDIYKNNKTTINENLIESKTDIKNNKATKEEKKKLYQAYDL
           STYKOLEAHNLISYLEKRIIDLTKNENIKELDKINELNKPANSGMTPNLLDK
           NKRIEHEKEPIKEIAKTIKFNIDSLFDPLELEYLKEKKNIDISAKVETKESTEN
           EYENGVTYPLSYNDINNALNELNSFDLINPDTKPSKNIYDNERKFIINKEK
           KIEKKKIESDKSIEDRSKSLNDITKEYKLINEIYDRYNNNIDLTNFKEMKRY
           SYVEKLTHTHTFASYSKSNLEKLTAKAKMWDYSLRNIYVEKELKYKKNLSKIE
           NEIETLVENIKRDEOLFEEKITRDEKRPBEKILEVDYKVOYOKYLANKIDELK
           TQLLKNVELKHNHVPNSTYKQENKOEPIYLIVLKEIDIKYFMPVESLINEEKN
           IKTEGQSDNSEPSTEGETIQOATTKPQOQSALEGSVOAQEOQAPVPVPVP
           EKAQOYTPPAPVNNKTEKNSKIDYLEKLEFLNTSYICHKYLIVSHSTNEKILKY
           KIRKEESKLSGDPDLDFNIONNIIPVMSMPSLNNLSLSOLMEIYKEMVCNLKY
           LKDNDRKIKNLIEKAKKYSTSVKTLSSSMGPLSLTPDQKDPVSNDDTSHSTNL
           KLFENILSLGKNNIYIELLGOKSSENFYEKILKDSFTFNESTNRYKSKADINSL
           NDESKRRLLEEDINKLTKTLQSLFDLNTKYLKERLPDKKTVGKMKOIKUTLLK
           EQLSKLNSLNNPKNHVNLFONFVFNKRNKEALATEMTLENTKILLHVGKLYVNG
           ESSPLKTLSEESIOTEDNVASILENFKVLKLEGLKDNLMLEKKISYLSGLHLILA
           ELKEVINKNTGNSPENNTDVNNALSESKKFLPEGTDAVTVVSESGDTLEBSOK
           KPSATHGASNTITTSQNDVEDVDIYIPIRESESDVDLGOVVTGMPVPSVD
           NIISIKENEVLYIKPLAGVYSIKOLNNMTEPVNNKTDILNSFNNRNNKNTL
           ESOLIPRKDLTSSNYVKDPYKELNKKRKPFLSSYIKDSIDTIDNPANDVLYIK
           ILSEKYSDDISIKYIINDQGENEKLPLFLNMLETYKYTNDKIDILFVHLEKAVLN
           YTYEKSNEVAKIKELANTYKTIODKLADPKNNNPNVGIADLSTDNHNNLLHLEKSTM
           VFENLAKTVLSNLDGNLQGLNLSOHCYKQCPQSGCFRHLDEDEEKCCLNTYKO
           EGDGCVENPPTCENNGGCDADAKCTEEDSGSKKTIQECRTPDSYPLFDGIFCS
           SNFLGISFLILMLILXSPF"
BASE COUNT 2363 a 608 c 676 g 1635 t
ORIGIN
Query Match      54.4%: Score 2686, 8; DB 3; Length 5282;
Best Local Similarity 71.8%: Pseq. No. 0;
Matches 3541; Conservative 0; Mismatches 1362; Indels 27; Gaps 1;

```

```

Db 405 GATGCTAATCTTACGCTGATTTAAACACAGAGTACGAATTAAGTATACAA 464
Qy 430 gagctgaagatcccaagctgttgcacctcaatcatatgctgcgaacgttgcatac 489
Db 465 GAACCAATATCCCAACCTTTGATTAATCATATCATATGATTAATCTTGTGATAT 524
Qy 490 attcagcttccaatcatctgattgagcgttaccagaagacgaatgaactccttacaag 549
Db 525 ATTCTGCTTTCAANTATTTAATTGATGATGATGAGAAATTAATGAATTTATATATA 584
Qy 550 ttgaattcctacttcgacttgcagggccaactgaaatgcgtttgcgccaatgactat 609
Db 585 TTTAACTTTATTTTGAATTTAATGAGCAAAATTAATGATGATGATGCAATGATAT 644
Qy 610 tgcgaatccatccaatltgaagatcgaagccaagcttgcagctatgaagaagtg 669
Db 645 TGTCAAAATCACTTCAACTTAAATTCGTCAATGATTAATGACGTACTTAAAAA 704
Qy 670 gtcttcgatatcgaagccctcgcagacaacatcaaggaacatgtgggaagaatgaag 729
Db 705 GTTTCGATATAGAAACCATTTAGACAAATTTAAAGTTAATGTAGGAAATGGAAGAT 764
Qy 730 tataataaagaataaagaagaccatcgaagacatlaacgaagctgatacgaatccaa 789
Db 765 TACATTTAAAAAATTAATAAATCCATAGAAAATTAATTAATGAATTAATGAAGAAGT 824
Qy 790 aagacatagacaataaataaagatgcaacccaagaagaagaagaagaagaagtgta 849
Db 825 AAAACAATGATTAATAAATAAATGCACTTAAACAGAGGAAAAAATTTATACCAA 884
Qy 850 gcccgatcgagccgctgcatacatalcaacaaagcttgaagaagccatacctatcag 909
Db 885 GCTCAATATGATCTTTTCAATTTACAAATTAACAAATTAAGAAGACAAATTAATTA 944
Qy 910 gtactggaagcgcatacagacccctcgaagaagaatgaataatcaagaagctgcag 969
Db 945 GTTTAGAAAACGATGACACCTTTAAAAAATGAATAACATTAAGATTAAGATTA 1004
Qy 970 aagatlaatgaatlaagaatccctccgcagccaactctggaagaacccctcaag 1029
Db 1005 AAGATTAATGAATTAATAAATCCCCACCGGCAATTCGTGAATTAACCCAAAT 1064
Qy 1030 ctggaacaagaacaagaatagagaagcagaagaagaatcaagaagatcccaaac 1089
Db 1065 CTTGATTAAGAACAAAAATGAGGACACGAAAAAGAAATTAAGAATTCGCAAACT 1124
Qy 1090 attaatcaacatagattcctcttacttactgaccccttgagctgagtgactactg 1149
Db 1125 ATTTAAATTAATATGATAGTTTATTTATCTGATCCACTGGAATTAAGATTAAGA 1184
Qy 1150 gagaagaataaataatagacatccgcgaagtcgaagacaagaagaatcaacgaact 1209
Db 1185 GAAAAAATTAATAAATATGATGATTAAGTCGCAAGGTTGAAAGAAATCACTGAA 1244
Qy 1210 aatgaatccccaatgtgtgacglaacccctgcttataacgaatcaacaagcctc 1269
Db 1245 AATGAATATCCAAATGAGATTAATCTTATCTTGTGATTAATGATTAATTAAGCT 1304
Qy 1270 aacgaagctcaatagcttcctgcgacttgaatlaacccctcgatlaacgaagaacct 1329
Db 1305 AATGAATTAATTTCTTTTGGATTAATTAATTAATCAATTTGATTAATTAAGAAC 1364
Qy 1330 aagaatatctacacagacaatgagagaagaagaattatcaacgaatcaagaagaag 1389
Db 1365 AAAAAATATATTAATGATTAATGAAGAAAAAATTAATTAATTAATGAAGAAAA 1424
Qy 1390 aaatltgagaagaagaagaattgagagtgacaagaagaagaatttcaacgaagccg 1449
Db 1425 AAAATAGAAAAAATAAATTAATGATTAATTAATTAATTAATTAATTAATTAAT 1484
Qy 1450 ctaaagatcatcaataaagaatgaaagaagctgtgaaagcgaagctatgattccaa 1509

```

Db 1485 TTAATGATATACAAAGATATGAAAAATTACTTAATGAAATTTATGATACAAATTC 1544
Qy 1510 aacaataacatcgacctgagcccaacttcgagaagaatgatggaacgctactctaca 1569
Db 1545 AATTAATATATATGATTTAACTAATTTGAAAAATATGATGGTAAAGATATTCATATMAA 1604
Qy 1570 gtaggaacacgacacacataatcctctgcaactctgagaatcttaagcaatctt 1629
Db 1605 GTTGAACAACTTACACACATATATCTTTGCACTTCTAAGAAATTCFAACATATCTTT 1664
Qy 1630 ggaagctacaagaagctcttaagatatgtgagagactatctctgagacactgtgtg 1689
Db 1665 GAAAGATTACAAAGCCTCTTAATATATGAGAGATTATTCCTTAAGAAATATGATGTT 1724
Qy 1690 gagaagaactaagaatlatcaagaatctcaatgaatgcgaacacgagatcgagacg 1749
Db 1725 GAAAAAGATTTAAATATTTATATAAATTTAATAGCAAAATAGAAAATGAGATTGAACA 1784
Qy 1750 ctgtctgagaacatlaagaagatgaagaacagctgtcttgagaagaagatlaacaagac 1809
Db 1785 TTAGTTGAAAAATATTTAAAAAGATGAAACACGCTTTTGAAAAAAAATTTACTAAAGAC 1844
Qy 1810 ggaataaaccagatgaagaagatcctgagagctccgatatgtttaaagctccaagtcag 1869
Db 1845 GAAATATTAACCGATGAAAAAATTTTAGAAGTATCTGACATTTGTAAAGTACAAAGTTCAA 1904
Qy 1870 aaggtctcctcatgaacagaatltgaactcaagaagaactcaactcatctcgaagac 1929
Db 1905 AAGGTTTATTTATGAAACAAATTTGACGAATTTAAAAAGCCTCAATTTGATTTAAAAAAT 1964
Qy 1930 gtggagcttaaaacataatatactatctgcccgaatgtctaaagcagaagaa 1989
Db 1965 GATGAAATTAACATATATATATACATGTCCCAATCTTACCAACCAAAAAATAAGCAAGAA 2024
Qy 1990 ccaatacctcatctgactcaagaagaagatagacacgaagctgtctccgacaa 2049
Db 2025 CCTTATATATTTAATTTGTGTTGAAAAAAGAAATGATTAATTTAAAGTTCATGCTTAAG 2084
Qy 2050 gtccgagagctctgaacagaagaagaacatlaaactgaagacagctcagaatac 2109
Db 2085 GTTAGATTCATTGTATTAATGAGAAAAAATCAATTAACAGAAAGCTCAATGATTAAT 2144
Qy 2110 tccgagctctccacagaagagagataaccgagcagctaccccaagcccgacaagc 2169
Db 2145 TCGGAACCTCAACCGAAGAGAAATACAGCAAGCAACATCAAAACCTGGACAAACA 2204
Qy 2170 gccggtctcagctcgcagacgaatgagctgcaagctcacaagaagcagaagcagagc 2229
Db 2205 GCAAGATCTGCTTTAAGAGAGATTCAGTACAGCAAGCAACAAACAAACAAACAAGCA 2264
Qy 2230 cagctccagctgcagtgccgcttccagaggtctaaagctcaagctcctacaaccagct 2289
Db 2265 CAACCCACAGTACACATACAGTACAGCAAGCAACAAAGTCCCAACACCAACAGCA 2324
Qy 2290 cctgtgaataaacaagccgaagaatgtcagcaaacctgactactctgagaagcctctg 2349
Db 2325 CCAGTAAATATTAATAAAGTAAAAATGTTCCAAATAGATTATCTTAAAAAATATATGA 2384
Qy 2350 ttcccttaatacactccacactctgcacacaatatatcctcgtctcacaagcacaagaac 2409
Db 2385 TTTTAAATACCTTCAATATATGTCACAAATATATTTGTTTACACACTCACTATGAAC 2444
Qy 2410 gagaagatctcttaacaagataaacaagaagaagaagagatlaaactgtcctctgt 2469
Db 2445 GAAAAATATTAATAAATATTAATAAATTAACAAGAGAGAAAGCAATTAAGTTCATG 2504
Qy 2470 gatccactgagactgtctgtcacaataccagaacaacatcccgctatctctatctgtc 2529
Db 2505 GATCCCTTATGACTTATGTTTATATATACAAATATACATACCTGTAATGTATGTATGTT 2564
Qy 2530 gatagctacaacatctctcctcactcagctgtlcatagagatatlaagaagagagctgc 2589
Db 2565 GATAGCTTAACATAGTATTATCACAACATATTATGGAATTTATGAAAAAGAAATGTT 2624

Qy 2590 tgaacctgtataaactcaagaacacgaacgaatlaagaacctctgtgaggagctlaag 2649
Db 2625 TGTAATTTATATAACTTATAGGATATATGACAAATTAATAAATTTATATAGGAAAGGAAA 2684
Qy 2650 aaggtctcacctctgtttaaactctctctccagctcctcagatcgaacacactgtctcaca 2709
Db 2685 AAGATTCACATCTGTATAAATCTTTCAAGTTCATCATCATCAACCATTAATCATTAACA 2744
Qy 2710 cctcaagaacaagccgaagtgaagcgtlaacgagacacctcctcaactcgaacacttaac 2769
Db 2745 CCTCAGATTAACCCGAATTAAGTGAATGATGATGATCATCTTCATTAATTTGAAT 2804
Qy 2770 aactcagaaactgtcttgaagaacatccgtctcctcgcgaagaataagaacatctacca 2829
Db 2805 AATAGTTTAAAAATTTATTTGAAAAACATATTGAGCTCTTGAAAAACAAAAATATATCCA 2864
Qy 2830 gaactatgtgacagaacatccgctccgagaaactctcagaagaagatctgaagaacgagc 2889
Db 2865 GAATTAATAGTCTCAAAAAAGTAAAGTAAACCTTTATGAAAAGATTAATAAAGATGCTGAT 2924
Qy 2890 acatctacaacgagagctcactcaactcgtgaaatcctcaagccgatgatacactc 2949
Db 2925 ACATTTATATATGAACTTTTACAAATTTGTATAATCTTAAGCTGATCATTTATTTA 2984
Qy 2950 cttacgatlgaatctaaacglaagaagctggaagagacatcaataagctgaagaagaca 3009
Db 2985 TTGAATGATGAATCAAAAAGAAAGAAATTTAGAAAGATTAATTAATTAATAAAAAAAT 3044
Qy 3010 ctgcaactgagcttgagactgtatacaagaagctacaactgaagaagagactctgcgc 3069
Db 3045 TTACAGTTTATCATTTGATTTATATATTAATTAATTAATTAATAAAGATTAATTTGAT 3104
Qy 3070 aagaagaagacagctcgcaagatlaagaatgcagatcaagaagctgtcctgcacagag 3129
Db 3105 AAAAAAGAAACAGTGTGTAATATATAAATGCAATTAATAAATCTTATTAATAAGAA 3164
Qy 3130 cagcttgaagcaacatcaactcaactcagaaatccgaacacgctactcagaactctca 3189
Db 3165 CAAATTGAATTCAAAAATTTGAATTCACATTAAACCCAAACAGTATTAACAAAACTTTCT 3224
Qy 3190 ggtctctcaagaagaagaagaagcgaagctgcgagacagaagaacactctgagaagac 3249
Db 3225 GTTTTCTTTAACAAAAAAGAGAGCTGAATATGCAAGAAATCGAANAACATTAGAAAC 3284
Qy 3250 accaagatctctcacaacactcaagaagcctcgtcaaglatataatgtgcagctctc 3309
Db 3285 ACAAAAATATTTATTAAGATTAATAAGCACTTTGTAATTAATTAATGCTGATCATCT 3344
Qy 3310 cctctgaagaactctcgcagagagagcatccgaacccgagataactaagccagctcgag 3369
Db 3345 CCATTTAAAACTTTAAGTGAAGATCAATTCAAACGAAAGATTAATATCCAGTTTAGAA 3404
Qy 3370 aactcaagctctcgtctcgaagcgaagcgaagctgaagaacaactcgaagcgaag 3429
Db 3405 AACTTTAAAGTATTAAGTAAATTAAGAAAGAAATTAAGGATTAATTAATTAAGAAAG 3464
Qy 3430 aagaagctcagctactcctcagcgaactgcacactcgtcgcgagctcaagaagac 3489
Db 3465 AAAAAATTAATCACTTATTAAGTGAAGTATTAATTAATTAATTAATTAATTAAGAAAGTA 3524
Qy 3490 atlaagaacaagaactacacgagcaatagcccaagcgaagaataacagcgtgaataac 3549
Db 3525 ATTAATAAATAAATAATATATCAGGTATTTCTCAAGTGAATAAATATACGATTTAAACAT 3584
Qy 3550 gaaactgaaacttacaagaagctcgtcgtcgaagaagaagcgtgcgactgtgtct 3609
Db 3585 GCATTTGAATCTTACAAAAAATTTCTCCAGAAAGAACAGATGTGCAACATTTGTAAGT 3644
Qy 3610 gaatcgtctcgaacacactgagcagctcacaactcaagaagcgtcactcactcagctgc 3669
Db 3645 GAAAGTGAATCCGACACATTTAGAACAAAGTCAACCAAAACCAACGACATCACTCATGTA 3704

OY	3670	ggggcggagctccaataacattaccacgtctcagaagcgtcgaagatgaggtgactaacgtc	37239
OY	3670		
Db	3705	ggagcagagcttAACAAATPACAAACATCCAAAAGTGCATGATCTCAAGTACATACGTAA	3764
OY	3730	atcattgacctactcttcgagcagagcgagaaagagactacatgacctgcgcgaagtgctc	3769
Db	3765		
OY	3790	accggtgagcgtcactcctctccgtgattgtataacattctgccccaaatctgagacgaa	3849
Db	3825	ACAGAGACAGACGATACCTCCCTCCGTAATGTATACATCTCTTCTTAAATGTAAATGAA	3884
OY	3850	tacgaatgctctactcgaacactctgcgcagcgctctatagctctctcaagaacaagctg	39039
Db	3885	TATGAGGTTTTATTTTAAACCTTTAGCAGCGTTTATAGAACTTTAAAAAACAAATTA	39444
OY	3910	gagataacggtgagacctccaatgtctaaacgttgaaagacattctgacacgcgccttat	3969
Db	3945	GAAATTAACGTTATGACATTTAATGTTAAAGTTAAAGGATATTTTAAATTACGATTTAAT	4004
OY	3970	aagagagaaatctcaagaaacgtctctggagcgactgattctccataaagacctgacc	40239
Db	4005	AAAGCTGAAATTTCAAAAATGTTTTAGATTCAGATTTAAATTCATATTAAGATTTTACA	4064
OY	4030	ctcccttaactacgtctgccaagagcccatataagttctcctcaataaagagaaggggataa	40839
Db	4065	TTCAAGTAAATTAATGTTGTCAAAGATCCATATAAATTTCTTAAATPAAAGAAAAAGAAATAAA	41244
OY	4090	ttctcgtctagttcaactactatcaagagctccatccgcacacgatalcaactctgctat	4149
Db	4125	TTCTTAAGCACTTTAATTTATTTAAAGATTCATATGATACGATATTAATTTTTCAAAT	41844
OY	4150	gactgctcgtgggtatttacaagatctctgagcgaaaaatacaagctctgacctgactctat	4209
Db	4185	GATGTTCTTGGAATTTATTAATAATTAATTAATTCGAAAAATTAATTAATTCAGATTTCAAT	42444
OY	4210	aaaaagatatcatcaacgcttaagcagagcggaatgaaataatctgcctctctgataac	42639
Db	4245	AAAAAATATATCAACGCAACAAAGGTAAGAAATGAAATAATACCTTCCCTTTTAAACAAT	4304
OY	4270	atcgaaacccctgtaacagagcagtgaaacgcaaaaaatcgacctcttgtaataacctgag	43229
Db	4305	ATTGAGACCTTATATATAACAGTTAATGATATAATTTGATTTATTGTATTTATTTAGAA	4364
OY	4330	gccaaagttcctcaactacttctcgaagaagcagatgtagagtttaaaatcaaaagagctg	4389
Db	4365	GCAAAAGTCTTAATTTATACATATGAGAAATCAAAAGTGAAGTTAAATTAABAACACTT	4424
OY	4390	aactactcctaaacatcatcaagacgaacgctgagcatctcaagaanaatacaaatctgct	4449
Db	4425	AATTAATCTTAAAAAACAAATTTCAAGCAAAATTTGGCATTTTAAAAAATAATTAACATTTCCCT	4484
OY	4450	ggaattcgaacccgtctaccgatttataacacaacaactctctgacaaagttctgtcc	4509
Db	4485	GGAAATTCGTATTTATACACAGATTTAATACCTAATAATTAATTTGACAAAAGTTCTTGT	4544
OY	4510	actggtacgtgtgtctcgaanaacccctgcgaanaacagtgctgagacatctgctcgaacgcac	4569
Db	4545	ACAGATATGTTTTTGAANAATCTTGTAANAACCGTTTTATCTAATTTACTTGATGGAAC	4604
OY	4570	ctgcaaggtacgtctgaacatctcccgacaccaaactcgtgaaagaacagtgccccagat	4629
Db	4605	TTTGCAAGGTATGTTTAAACATTTTCAACAAACCAATGCGTAAAAAACAATGTCACAAAT	4664
OY	4630	agcgccgtcttcagacatctgcgcagcgccggaagatgaaatgtgtctctcgtgaactcaaa	4689
Db	4665	TTCTGATGTTTTCAACACTTTAGATGAAAGAGAGATGTAAATGTTTTATTAAATTTACAA	4724
OY	4690	caagaagagataagtgctggtgagaaccccaaaccttactcgtcaatgaaacatctgcgg	4749
Db	4725	CAAGAGGTATATAATGTTGTTGAANAATCCAAATCCTACTCTGTATACGAAATATATGTGGA	4784
OY	4750	tgtgacgcgcgattgctaaatgcaccgcgaagacagcgcgtctaacggaagaanaatcaca	4809

Db.	4785	TGTGATGCCAGATGCCAAATGTACCCAGACAGATTAAGGTAGCAGACGAAAGAAAATTCACA	4844
OY	4810	tgcagatctactaagcccgcagctcctatcaacatcttcgcagggatttttgcgccagctc	4869
Db	4845	TGTGATGTACTAAACCGATCTTATTCACATCTTTGCATGTGATTTTTCGCGATTCCTCT	4904
OY	4870	aattccctgggacatctcctctcgtcatgcatacgtatgcctgtacagcttactaa	4929
Db	4905	AACCTCTTAGGAATTCATCTTATTATTAATACATGTTAAATATATACAGTTTCATTTAA	4964
OY	4930	tagatcatgt	4939
Db	4965	AAAATGTAGG	4974
RESULT	9		
LOCUS	AF062349	5271 bp	DNA linear INV 09-FEB-2001
DEFINITION	Plasmodium falciparum strain HN2 merozoite surface protein 1		
ACCESSION	AF062349		
VERSION	AF062349.1		
KEYWORDS	GI:3859844		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota, Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5271)		
TITLE	Jiang, G., Liu, R., Daubenberger, C. A. and Pluschke, G.		
JOURNAL	Sequence analysis of the MSP 1 gene of Plasmodium falciparum from		
REFERENCE	Hainan, China		
AUTHORS	Chung, Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17		
TITLE	(5), 294-297 (1999)		
JOURNAL	2 (bases 1 to 5271)		
REFERENCE	Jiang, G., Liu, R. Z., Daubenberger, C. A. and Pluschke, G.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical		
REFERENCE	Institute, Socinstrasse 57, Basel CH-4002, Switzerland		
TITLE	Location/Qualifiers		
FEATURES	1..5271		
SOURCE	/organism="Plasmodium falciparum"		
	/strain="HN2"		
	/db_xref="taxon:5833"		
	/chromosome="9"		
	/country="China"		
	<66..>5180		
	/gene="msp1"		
	<66..>5180		
	/product="merozoite surface protein 1 precursor"		
	<66..>5180		
	/gene="msp1"		
	<66..>5180		
	/gene="msp1"		
	/note="binds spectrin"		
	/codon_start=1		
	/product="merozoite surface protein 1 precursor"		
	/protein_id="AAC72885.1"		
	/db_xref="GI:3859845"		
	<66..>5180		
	/translation="MKIIIFPLCSFLPIINTOCYTHESYOELVLEALDAEVALTGYR		
	LFGKSNRTNCTGSTAVTSTPGSGSGGSAVGSAGSVASVAGSVASVAVS		
	GGGNSKRTNCTNDSGSDASVADLKHRYNYLFTIKELYPPLPULTNMLLCPDN		
	IGHKRYLDIGEEINELLYDKNFYDILRAKLNDACASTQOIFNPLIRNELDVLK		
	KIVAGYRRPLDNIKNDVGMEDYIKKKTTITANLEIEGSKITDKNANBNBEGK		
	KLYAOYNLFIYNKQLOEAMNLSVLEKRIPTLAKNEIKKLIDLEIKNPANGSG		
	NTPTLIDKNKIEHEBEKIKELAKTIFNDISDLEIYLRKNNKVDVTPKSGD		
	ODPKSVQIPKVPNGVIVPLDITDHNLSAANDKNSYDGLLELNNPTKKEINKIT		
	DNKRKRFINNIKQIDLEENINHTKEQKKLLEDYEKSKDYEBELFEREKKPN		
	NFDDVADVKITFSARYTYNVEKQRYNNKPFSSNSNVYVQIKKALSTLEDLSKKGIS		
	EKDNDNHYTLTKGLEADIKKLTEIKSSENKILKNNKGLTHSNASLESDIVKLOV		
	OKVLLIKKIDELRKIELEFLKNAOKDSIHVNIYKPNKPEPYLLIVLKKVEDLK		
	IPKDKMLKKKQAVLSITQPIVAVASFTEDGSGSTHLSQSGEYVTEFEKEE		
	GHTTIVTITLPTPTQSPPEKYYVENSIEKSDNSQALIKTYVILKLDPEIKSYIC		
	HKYLVLSNNSMDKULEVYNLTPEKNEKLSCDPLDLFNPQNNIPAMYSLYDSMD		
	LQHLFELVYOKEMIVYLHKLEENHIKKLLEEQKQITGSTSTSGPNTVTAQSAVH		

SNSQNOQNASSTNTONGVAVSSGPAVVEBSHDLPTVLVSTISNDLKGIVSLNLGNKTK
VNPPLTISTTEMEKEFYENILKKNNDYFENDIDKOFKNSKRVITGLLETQKMLNDEIK
KLKDLTLOSPDLFNYKXKILDLFLFNKKGODKQKINKLTLTGLLETSKLSLNPH
NVLONFSPFNKKEAEIETENTLENTLILKHKGALKYNNKESPLKTTSEVSIQ
TENDYANLEKFPVLKIDGKLDNLHLGKLSLITSLSSGLHHITELKEVTKNTGN
SPSENNKRYNEALKSYLENLPKAKYTYVTPPODYTPSPLSVRVSSGSKTEEOI
PTSGSLTELOOVOSONTDEEDSLVLYLPIGESDNDYEDQVVTGALISVMDNI
LSGFENERVDYILKPLAGVRSJKQIEKNITTFNLNLDILNSRLKRRKFLDYLES
DLMOFKHLSNNEYIIEDSPKLSNDEOKNLKSYKIKXSVEENDIKFAOEGISYEKV
LAKYKDDLESIKKIVEKEKPESSPTTPSPKDEOKKESKDFLPLNITLIVNLI
VKNIDYILNLAKIKINDCNVEDEAHVKTITKSLDKAIDKIDLEFNTNDFPAIKKLI
NDDTKRDMIGKLSLSTKINDONPENTTISIKRFPDMLNISOHOCYKRCQPCNSCER
HIDDERECKCLINIKOEBDKCYENPMPTNENNGGCDADAKCTEEDSGSNKKTJCEC
TKPDSYPLFDGIFCSSSNFLGISFLILMLILYSFI"

BASE COUNT 2313 a 687 c 690 g 1581 t
ORIGIN

Query Match 28.4%; Score 1400.6; DB 3; Length 5271;
Best Local Similarity 57.3%; Pred. No. 5.4e-298;
Matches 2944; Conservative 0; Mismatches 1974; Indels 219; Gaps 14;

QY 10 atfaaaatcattcttcctcgttcaattctglttltatcaatcaatcagtgctg 69
DB 66 ATGAAGATCATATCTTTTATGTCATTTCTTTTATTAATAATACACATATGTA 125
QY 70 acccaagaatccatcaagaagcgtgtaagaaccttgagaagatgcgcctcct 129
DB 126 ACACATGAAGAAGTTATCAAGAACTTGCAAAAACTTAGAAGATGCAATATG 185
QY 130 accgaatacagcctgttccagaagaagaatgctgtaagaaggagcagtgcaag 189
DB 186 ACAGGTTATGTTTATTTCAAGAAAAAATGATTAATAGAGAACAGTGGACAA 245
QY 190 gcaagttacaacacagcacaccggcttcaagggtctgtgctagcagtgctgctg 249
DB 246 GCTGTTACAACTAGTACACTGCTTCAAGGTTCAAGGTGCTTCAAGGTGCT 305
QY 250 tctgtgacctgtgagggttccgtcgcctccggcgagcagtgagcatcagtgctcagtg 309
DB 306 TCAGGTGTTGTCAGGTGCTGCTTCAAGTTGCTTCAAGGTGCTGCTGCTGCT 365
QY 310 gcaagcggtgctcgggaacagtcgaagaacatccatctgacacatctagcgtatcc 369
DB 366 GCTTCAGGTGCTTCAGTAATTCAGACGTACAATTCCTTCAGATTAATTCAGATTC 425
QY 370 gaagcaagaagcctcagcgcctcaagaacggagtgagaacatctccatcatcaag 429
DB 426 GATGCTAAATCTTACGCTGATTTAAACATAGAGTTCAAAATTAATCTTCACTATTA 485
QY 430 gaagctgaagtaaccacagtgctgaacctcaatcatatgctgacacatgctgataac 489
DB 486 GAACCTAAATATCCGAACTCTTGATTTAACCAATCATATGTTAATCTTGTGATAT 545
QY 490 attcaatgagcttaaatatcgtatgacggttaagaagagatcaatgaccccgtaacag 549
DB 546 ATTCAATGCTTCAATATTTAATGATGATGATGAAGAAATTAATGATTAATTAATA 605
QY 550 ttgaattctactcagctgtcgaaggccaacgtgaacgtttgccaatgagat 609
DB 606 TTTAACTTTATATGATTTATTAAGACCAATTAATATGATGATGCTGCAATAGTAT 665
QY 610 tgcacaatccatccaattgaagatcagaacgaagcttgagcgtatgaagaagtg 669
DB 666 TGTCAATACCTTTCAATCTTTAAATTCGTCAAAATGATAGCGTAAATAAAAAAT 725
QY 670 gtctctcgatatacgcaagcctccgacaacatcaaggacatgtgggaagaatggaat 729
DB 726 GTGTTGGATATGAAAAACCATTAAGCAATTAATTAAGATATGATAGAAAAAGGAAG 785
QY 730 tatataaagaataagaagacatcagaacatgaagcgtgacgaagaatccaa 789
DB 786 TACATTTAAAAAATAAACAACCATTAAGCAATTAATTAATTAATTTGAAGAGATAG 845

QY 790 aagacacatagacacaaaaataagaatgcaaccaagggagaagaagaagtgataccag 849
DB 846 AAAACATTTGATCAAAATAAAGATGAGATTAATGAAGAAGAAAAAATAATTAACAA 905
QY 850 gcccaatagacactgtccatctatacaaaacagcgtgaagaagcccaataactcaacag 909
DB 906 GCTCAATATTAATCTTTTATTTAATTAATTAATTAATTAATTAATTAATTAATTA 965
QY 910 gtctcgagaagagcgtatagacaccccaagaagaatgaaataccaagaactgcgcag 969
DB 966 GTTTTGAAGAAAAAGTATGACCTTTTAAAAAATAAAGCAATTAAGAAATTAATCTGAT 1025
QY 970 aagatataagaatlaagaatcctccgcaagccaactctggaacaacccctaaccgtg 1029
DB 1026 AAGATTAATGAATTTAAAAATCCCGCCCAATTCGTGAATTAACCAATTAATCTCTC 1085
QY 1030 ctggaacaagaacaagaatagagagcgcgagaagaagatcaagaagatcgccaac 1089
DB 1086 CTGTGATTAAGAAACAAAAATCGAGAACACAGAAAGAAAAATAAATAATTCGCAAACT 1145
QY 1090 attaaatcaacatagatctcctcttactgactcccttgagcgtgagtaactcattgaga 1149
DB 1146 ATTAATTTTATACATTTGATGATTTATTTACTGATCAGCTTGATTAATTAATTTA 1205
QY 1150 gagaagaataagaataatagacatctccgcaaaagtcga-----gacaagaatca 1200
DB 1206 GAAAAAATAATTAATAAGTGTATGATACCTTAATTCACAAAGATCCTGCAAAATCTGTCA 1265
QY 1201 accgaacctaatgaatatcccaatggtgtgagcgtaccctctgtctataagatataac 1260
DB 1266 ATACCAAAAGTCTTATCAAAATGATTAATTTATTTACCTACCTGCTGATTTCAAT 1325
QY 1261 aa-----cgctccaacgagcctcaatagctcgtgactgataaaccctcgat 1311
DB 1326 AATTCTATGACCTGACATATATGATTAATAATTCATATGATGATTAATTAATCTGAT 1385
QY 1312 tataagaagaacccctctaagaatctatacagaacaatgagaagaagttatcaac 1371
DB 1386 AAAAAAATAATTAATGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1445
QY 1372 gaatacagaagaagaatcaaaatltgagaagaagaagaatgagatgacagaagaatltac 1431
DB 1446 AACATTTAAAAACAATAATTTGATTTAGAAAGAAAAACAATTAATTAATTAATTAAT 1505
QY 1432 gaagaccgcagcaaaagctcaaacgaatatacctaagaaglatgaaagctgctgaagag 1491
DB 1506 AAAAAATTTACTGGAAGATTAATGAAGAGTCAAAAAAGAGATTAATGAAGATTTG 1565
QY 1492 atctatgattccaatccaataaataacatcgaccctgacccaactcgagaagaatgagga 1551
DB 1566 TTTTATGAATAATGAATAATTTAATAATTAATTTGACAAAGATGCTGTAGATTAATAAT 1625
QY 1552 aaacggtaactctacaagaagtgaggaagcgtgacacacataatcctttcatcctatgag 1611
DB 1626 GCAAGATATATCATATTAATGTTGAAAAACAAGATTAATTAATTAATTAATTTTCA 1685
QY 1612 aattcagaacataatcttgaagaagctcaccaaaagccttcaaglatatggagagacttct 1671
DB 1686 AATTCTGATATTAATGTTTCAAAAAATTAATAAAGAGCTTTCATATCTTGAAGATTA 1745
QY 1672 ctgaggaacatgtgtgtggaagaagaactaaaglatcaagaatctcataagtaagac 1731
DB 1746 TTTAAGAAAAAGAAATTTCTGAAAAAGATTTTAAATCATTAATTAATTTGAAAACT 1805
QY 1732 gaaaacgagatcgagcgcgtgtgtgagaacatlaagaagaatgagaagaagtggttgg 1791
DB 1806 GAAGCTGATTAATTAATAAATAATTAACAGAAAGATTAAGATGATGAATAAATAAT 1865
QY 1792 aagaagattacaaaagaagcaaaataaaccagatgagaagaatcctcgagggtctccga 1851
DB 1866 AAAAAATTTTAAGACTTAACACATTTGAGCAAAAG-----CTTCCTTAAGAGATTA 1922


```
Db 4077 GGAGTATATAGAGCTTAAAAAACAATTGAAAAAACATTAATTACATTTAAATTTAAAT 4136
Oy 3940 gtaagagacatctctgaacagcgctttaataagagaanaattccaagaacgctctggag 3999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4137 TTGACAGATATCTTAAATTACGCTTTAGCAAAACGAAAAATTTCTTAGATGATTAGAA 4196
Oy 4000 agcgacttgatccctataaagacctgacctcctctaactagctgltcaagaccatc 4059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4197 TCTGATTTAATGCAATTTAAACATATATCTCAAAATGAATACATTTATTGAAGATTCATT 4256
Oy 4060 aagttcctcaataaagaagaagagataaattctgtctagtctatacactatacagaagac 4119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4257 AAATTTATGTAATTCAGAACAAAAACACCTTTTAAAAAGTTTAAATATATATATATATAT 4316
Oy 4120 tccatcgacaccgatatcaattcgtctaagatgctggggattacaagatccctgagc 4179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4317 TCAGTATGAAAAATATATTTAATTTGCACAGAGATATAGTTATTTATGAAAAGTTTAA 4376
Oy 4180 gaaaaatacaagtcgaccttgacctctatataaagaatatacaagataag----- 4230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4377 GCGAAATATATAGAGATGATTTAGATCATATTAATAAAAGTTATCAAGAGAAAGAGATTTC 4436
Oy 4231 -----caagcgagaaat 4242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4437 CCATCATCACCCACCAACACACCTCCGTCACCAAGCAAAAAAGACGAAAGAAAGAGAA 4496
Oy 4243 gaaaaatcctgcctcctcctcgaataacatcgaacccctgacagaacgtaacgacaa 4302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4497 AGTAAAGTCTCTCATTTTAAACAAACATTGAGACCTTTATACATTAATCTTAGTTAATAA 4556
Oy 4303 atcgacctctgtaattcaccctggagagccaaagctcctaactatacttaagaaagagc 4362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4557 ATTGACGATTAATCTTAATTAATTAAGGCAAAAGATTAAACATTTGTAATGTTGAAAAA 4616
Oy 4363 aatgltgaagatlaaatacaagagctgaactccaacaacatccaagaacgctgca 4422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4617 GAAGCAACATGTTAAATTAACCTTAAGCTTAGATTTAAACCAATGATGCAACAAATAGAT 4676
Oy 4423 gatttcaagaanaatacaaatctcgtcggaattcgacagacgtctctacgataaacac 4482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4677 CTTTTAAAAAACATCAAGCACTTCGAAGCAATTAAAAAATGTAATGATGATACGAAA 4736
Oy 4483 aaaaactcctgcgaagtctcgtcacacgcatggttgcgaanaacctgcgaacaa 4542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4737 AAAGATATGCTTGCAATTAATTAAGTACAGATTAAG--TTCAAAATTTTCTTAATACA 4793
Oy 4543 gtcgtgaagacatctgcgcgaacacctgcagagcactgacgaacatctccacgaacaa 4602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4794 ATATATATCAAAATTAATTAAGAGAAATTCACAGATATGTTAAACATTTCAACAAACCA 4853
Oy 4603 tgcgtgaagaagaacagtgccccaagaatagcgctgtttcaagcacttgcagagcgcgaa 4662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4854 TGCCTAAAAAACAATGTCAGAAAAATTCGAGATGTTTTCAGACATTTAGATGAAGAGAA 4913
Oy 4663 gaatgaagaagctcctcgaactcaaaagaagaagataagtgctgagaaacccaac 4722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4914 GAATGTAAATGTTTATTAATTAACAAACAGAGGTGATTAATGTGTAATAATCCAAAT 4973
Oy 4723 cccaactgaatgaacaaatgcgcgagtgtaagccgacttaaatcaccgagagagac 4782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4974 CCACTTTTGAACGAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5033
Oy 4783 agcgactcaagcgaagaagaatacaatcagtgagtgtaactaagcccgactctatacactc 4842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5034 TCAGGTAGACCAACGAAAGAAATACACATGTGAATGTACTTAACCTGATTTCTTATCCACT 5093
Oy 4843 ttgagaggaatttttgcctcgaactcaatttctgagacatccctcctcgctgactcctc 4902
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5094 TTGAGATGATATTTTCTGCACTTCTTAACCTTCTTGAAGATATCATTTTAATTAATCTC 5153
Oy 4903 atgctgactctgaagcctcactcaataagatcagat 4939
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5154 ATGTTAATATTTATACAGCTTCAATTTAAAAAAATGATAG 5190
```

RESULT	10	AF062348	5243 bp	DNA	linear	INV 09-FEB-2001
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						

OY	4129	accgatcaatcttcgcgaatgagctgctgggtatctcaagatctcgaagcaaaataac	4188
Db	4296	AATGATATTAAATTTGGCACGAGAGGTAAAGTTATTTGAAAAAGTTTAAAGCGAAATAT	4355
OY	4189	aagctgcagcttgactctataaagaatatacaacgaataag-----	4230
Db	4356	AAGGATGATTTTAGATTCATTTAAAAAGTTTATCAAGAAAGAAAGAGCTTCCATCATCA	4415
OY	4231	-----caagcgagaatgaaaaatac	4251
Db	4416	CCACCACACACCTCCCTCACACACAAAACAGACGAGACAAAAGACGAAAGATTAAGTTC	4475
OY	4252	ctgcacctcgaataaacatcgaaacccctgtgtaacagacagtgaaacgacaaatcgacctc	4311
Db	4476	CTTCATTTTTRACAAACATTTAGAGCTTTATTCATTAACCTTAGTTAATTAATTTAGCAT	4535
OY	4312	ctcgtaattcaccttgagccaaggtccctcaactaactacacgaagaagcaatgtgaa	4371
Db	4536	TACTTAATTAATCTTAAAGCAAAAGATTAAACGATTTGATTTGTTGAAAAAGATGAACACAT	4595
OY	4372	gtlaaataccaagagctgaactactactaaacaataccaagaacagcttgcagattcaag	4431
Db	4596	GTTAAATAACTAAACCTTAGATTTTAAAGCAATGTGATGACAAATATGATCTTTTAA	4655
OY	4432	aaaaatacaaatcttcgtcggaatgcagacccgtctaccgaattataaccacaacatctc	4491
Db	4656	AACACTAAGCATTCGAGAGCATTTAAAAAATTGATTAATGATGATGACGAAAAAAGATATG	4715
OY	4492	ctgaccaaagctctgtccacttgcacttgcagtgtgttcgaaaaacctcgccaacagctgcagc	4551
Db	4716	CTTGGCAAAATTAATTAGTACAGAGATTAG---TTCAAAAATTTTCCATTAATTAATTAATCA	4772
OY	4552	aatcgcccgagcgaacctgcgaaggtcactgtcgaaacatctcccgaccacaaatgcgtag	4611
Db	4773	AAATTAATTGAAGGAAATTTCCAAATATGTTAAACATTTCCACACACACATGCGTTAAA	4832
OY	4612	aaacagtgccccagaataagcggctgttcttcaggcaactgcgacgagcgcgaagatgcgaag	4671
Db	4833	AAACATATCTCCAGAAAAATTCGGAGTTTTCAGACATTTAGATGAAAGGAAGAAATGTAAA	4892
OY	4672	tgtccctcgaaactacaacaagaagagataagctgcgtgggaagccaaacctcctgc	4731
Db	4893	TGTTATTATTAATTCACAAACAGAGAGGTATTAATGTGTGAAAAATCCAAATCTCTACTTGT	4952
OY	4732	aatgaanaacaatgycgggtgtgacgcgcgaactaaatgacacgaggaagacagcgctct	4791
Db	4953	AACGAAATTAATGTGGGTGATGTGATGCACATGCCCCAATATACCGAAGAGAAATTCAGGTAGC	5012
OY	4792	aaggaagaagaatacacatgagagtgactgaagcccgactctctacgctctcttcgacggg	4851
Db	5013	AACGAAAGAAATACATGTGAATGTCTAAACCTGTATCTTATACCACTTTTGCATGTGT	5072
OY	4852	atttttctccagctcacaattcctctgggcaactcctccctgcgtatcattcattgcgtac	4911
Db	5073	ATTTCCTGCAGTCTCTAACTTCTTAGAGAAATATCATTTCTTAATTAATCATGTTAATA	5132
OY	4912	ctgtacagctcattcattatagatcgatg	4939
Db	5133	TTATACAGTTTCAATTTAAAAAATGTAGG	5160

RESULT	11
LOCUS	AF218248
DEFINITION	AF218248 5085 bp DNA linear INV 24-JAN-2000 Plasmodium falciparum major merozoite surface antigen (gp195) gene, complete cds.
ACCESSION	AF218248
VERSION	AF218248.1 GI:6740099
KEYWORDS	
SOURCE	malaria parasite P. falciparum.
ORGANISM	Plasmodium falciparum
REFERENCE	Eunaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 5085)

AUTHORS	Shan, Z.X., Yu, X.B., Li, X.R., Ma, C.L. and Fang, J.M.
TITLE	Molecular cloning and sequence analysis of major merozoite surface antigen(gp195) gene of <i>Plasmodium falciparum</i> isolate FCCI/HN
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 5085)
AUTHORS	Shan, Z.X.
TITLE	Direct Submission
JOURNAL	Submitted (21-DEC-1999) Department of Parasitology, Sun Yat-sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, P.R.China
FEATURES	Location/Qualifiers

```

URES      location/Qualifiers
source    1. .5085
           /organism="Plasmodium falciparum"
           /isolate="FCC1/HN"
           /db_xref="taxon:5833"
           /country="China"
gene      1. .5085
           /gene="gpi95"
DSD       1. .5085

```

/product="maior mercaptoic surface antigen"
 /protein_id="AA027526.1"
 /db_xref="GI:6740100"
 /translation="MKIIIFLCSLPLFIINTQCTVHSELYELVKLEALDVAITGV
 LPOKEMVINEGTSCTAAVTTPSGSGSVTSGSGGSAVASAGSGSGVASGSGG
 RRPNSDNGSDSDAKSYADILKRVONTYITIKELYPELDTLNNHMLTLDNTHGFEF
 LIDYGEINIELLYKINFPYDILIRAKINAGANSYCOIPEMLKIRANLVDLKKIVGE
 RPLDINKQNGKEMEDYIKKNTYTIANNEIDSKTITIDONKNADEBKRLKXOR
 YNLFIYKNQJQEAHNLISYLEKRIIDLTKNENIKKLEDDIKITDAEKITTSKXPY
 LPENNKKEVEGHEEKIKELATIKTINDISLTDPLELEYIRKNNKVDVTPKSODP
 KSVQJLEPPYNGVILPLDILIHNSLDAADKSYDQDLNMPDKETINKIITDNK
 RKLEIFNNIKQIDLEEKINHTGEQNKRLDNDYKSKQDVLEPEKKEENKFPNNEP
 DAVDQITFSARATYVNEKQYNNK FSSNSYVYNQAKLKSALSYEDSLKRGISEKD
 NHYITLDTGLADIKKLEKEIKESSENKILEKNFGDLHSAASLEVDYKLOVOKY
 LKIKEDDKRIELPLKNAQKQIHPVPIYRPNKPEPVIYLVKREVDKLEIFIR
 KMKKKEADAVLISSTOPLVAASETTEGGSHTHLQSGEETEEVYVGGHTVITLI
 PPOPSPREKVEVENSIEHKSNNQSALITVYVILKKLDELPTKSYICXKIIYNSNS
 MODKLEIYVNLPEKNEKLSQDPLDILFIQNNIIPMTSLITSDMANLOHLEFEL
 KEMITVLIHKLEENHIKKLEBOKQITGTSSTSGNTVYVTAASHNSQMOQSS
 SYNTDNGVAVSSGPAVVEESHDPLTVTISINDKGLVITSLGNKKNRPNPLTIS
 EMKEFTEGINKNDYDIPNDIDIKQFYSKSVITKLTETQKALNDEIKKLDITLQTS
 DLYNKKIKLKLDTLPKKKKEGODMKQIKVILKLEQLEESKLSNSLNHNANLQNSVFL
 NKKKEELIETENTLENTLRKILKHGKQILVYKYNLEESSPLKLTSEVSIQTEPDANILE
 FVYLSKTODKLNHLHKKKKSPLSSGLNHLITELKVEYKNNKYTQNSSENNKRYD
 EAKSTENLPEPAKVTYVTPPODVTPEPLSVYSSSGSTKRETIPIPGSLILIE
 QOOWSONYDEEDSDSLVVLPIGSEBNDVLDQVYGEALISVMDNIIISGENEYD
 IYKPLAGVYRSLKQIEKNIITFNLNIDNLSNLRKKRYDILVIESDLMQKRYI
 NEYIIIEESFRLNSLOKQNTLKSQYIIEESYENDIKFAQESIISEYEVYAKYADDI
 IKKVIIEKEKEPSPPTPPSPAPKIDDECKKIDKIDTLPKNTPEALIKKLIINDTKDMH
 LKAKINDCWEKDEAHVAKITKISDLKAIKDKIDILFKMNTPEALIKKLIINDTKDMH
 KLSLQVONPNTIISKLIBSKRODMANISHOQYVKQCPENSGCRHHLDEBECK
 LMKVQBGDKVYENPPTCENENGGCDADACTEEDSGNKKKITTCTYPSDIPULF
 GIISSCSNLGISLITLMLILLYST"

Query Match	26.9%	Score 1330.8	DB 3	Length 5085
Best Local Similarity	57.0%	Pred. No. 1.3e+282		
Matches 2915	Conservative 0	Mismatches 1972	Indels 231	Gaps 17
Qy	10	atgaagaatcattcttcctcctgcgttcattccgtttttatcatacatcagagcgtg	69	
Db	1	ATGAGAGATCATATCTTTTATCTTCATTTCTTTTATATATATCAACATGTCTA	60	
Qy	70	accacgaatccatcagagagcgtgtaagaacatcgaaagccttggaaagatgcgcctt	129	
Db	61	ACAAATGAAGTTATCAAGAACCTTGTCAAAACAACTTGAAGAGTGAAGCATGTATG	120	
Qy	130	accggaatacagcctgttccagaagaagaatgctgtaatgaagaaggacagatggcagc	189	
Db	121	ACAGCTATATGTTTATTTCAAAGCAAAAATATGATTTAAATGAAGCAACAACTGGAAACA	180	

OY	190	gqcglttaacacagcaacacccggtlctaaaggctcgtgtcgaagcggtgcgcggttg	245
Db	181	gctgttttcaactatctaacacctgtgtcaggtgcgttccagttccaggttcacgtt	240
OY	250	tcctgtgcctcttgagggttcgcgcgcctccgcggcagcgctgcgaatcaggtgcacgtg	309
Db	241	TCAGTTGCTTCAGTTGCTTC-----AGGTGGTTCCAGTGGCTCACTT	282
OY	310	gcaagcgcggtlcccggaacagltcgaaagacatccatctgaacactlagcgatcc	369
Db	283	GCTTCAGGTGGTTGAGTAAATTCGAAGCAGTACAAATCTTCGATGAATTAATCAAGTATCA	342
OY	370	gaagccaaagtcctcagccgcacccaagcagcaggtgagaaactatctcctaatacaag	429
Db	343	GATGCTAAATCTTCGCTGATTTAAACATAGAGTCCAAATTAATCTGTCTCTATTA	402
OY	430	gagctgaagtaacccaagctgttcgcacctcaactacatctgtcacactgtgtatac	489
Db	403	GAACTCAAAATTCGCCAAGCTTTTGATTTAACCAATCATATGTAACTTTGGTGATAT	462
OY	490	atcaatgcctcaaatctcgtatctgattgcggttlacgaagagatcaatgaactccgtlacaag	549
Db	463	ATTGATGGTTCCAAATTTTAATGTGATGATAGAAAGAAATTAAGCAATTAATTAATTA	522
OY	550	ctgaatttcctcactcgcgactctgcgaaggccaaactgaatgaagtttcgccaatgaacat	609
Db	523	TTAAACTTTTATTTGATTTATTAGAGCCAAATTAATGATGATGTCATAGTAT	582
OY	610	tgtaaatlccatccaatlctgaagatcagaagccaaagcgtltgacgtaltgaagaagltg	669
Db	583	TGTCAAAATACCTTTCAATCTTAAATTCGTCGCAATAGATTTAGAGTAAATTA	642
OY	670	gtcttcggatctgcgaagcctctcgcaaacatcaagacaatgtytggaagatgaagat	729
Db	643	GTGTTCCGATATTAACAAACCATTAAGACAAATTTAAAGATTAATGTGGAATAAT	702
OY	730	tatattaaagaagtaagaagccatctgaaacatttaagagctgatcgaagaatccaa	789
Db	703	TACATTTAAAAAATTAACACCATTAAGCAATTAATTAATGATTTATTCAAGAGATG	762
OY	790	aagaccatagacaanaaataagaaatgcacaacgaaggaagaagaagaagtlgtcacag	849
Db	763	AAAACAATTCATCAAAATTAAGAAATGCGAGATTAAGAGAAAGAAAAAATTAATACAA	822
OY	850	gcccaagtaagcctgtccatctataacaacagcttgaagaagcccatcaactatcgc	909
Db	823	GCTCAATTAATCTTTTATTATTAACAAATTAACAAATTCGAAGACACATTAATTAATAGC	882
OY	910	gtactcgaagaagcgcatagagcaccctcaagaagaatgaaatatacaagaactgtgtcac	969
Db	883	GTTTTAGAAAAACGTATTGACACTTTTAAAAAAAATGAAAACATTAAGAAATTAATCTTAA	942
OY	970	aagatlaatgaatlaaga--atcctccgcgaacccaactcttgggaacaccccctlacag	1026
Db	943	GATATAGATTAATAATTAACAGATGCGGAAAACTCAGTATGGAAGTAAACAAATCTT	1002
OY	1027	ctgtctgagcaagaacagaagatag---agagacacgaagaagaatcaagaagatgc	1083
Db	1003	CTCCCTGGAATTAAGAAAAAGAAAGTCGAGGACACGAAGAAAAAATAAAGAAATTTCC	1062
OY	1084	aaaaccatlaagttcaacatagatctctcttlaacgatccctctgaagcttgagtaacac	1143
Db	1063	AAAACATTAATTAATTAACATTTGATAGTTTATTACTGATCCACTTGAATTAATTAAT	1122
OY	1144	ctgagagagaagaaatgaatatagacatctccgcgaagtcga-----gacaag	1194
Db	1123	TTAAGAGAAAAAATTAAGAAAGTTGATGTAACACCTAAATTCACAAAGATCTTCGAAATCT	1182
OY	1195	gaatcaaccggaactaatgaatatcccaatggtgtgacgtacccctctgtctataagat	1254
Db	1183	GTTCAAATACCAAAAGTTCTTATTCCAAAATGATGTATATCTTTACCACTCACTGAT	1242

QY	1255	atcaacaa-----cgcttcaacagcctaaatagcttggtgagcttgattaaacct	1305
Db	1243	ATTGATATTTCATTAGCTGCAGATTAATGATMAAAATTCATFAGTGATTTAATGAATCT	1302
QY	1306	ttcgattatacgaagaacctctaagaatatactacacagacaatfagagaagaagctt	1365
Db	1303	GATFCTAAAGAAAAAATTAAATGAAAAAATTATTCAGATTAATTAAGAAAAAATAATTC	1362
QY	1366	atcaacgaatcaagsgaagatcaaaattgagaagaagaaattgagatgcagaaga	1425
Db	1363	ATTATATAACATTAATAAAAAACAAATTTGATTGAGAAAAAACAATTATACACAAAAAGAA	1422
QY	1426	agttacgaagacccgcagcaaaagcttcaacgatatacctcaagaagatfagaagaagctgct	1485
Db	1423	CAAAATTAATAAATTTACTTGGAAAGATTATGAAAGTCAAAAAAGATTTTGAAAGATTTACTT	1482
QY	1486	aacgagatctatgattccaattcaacaataacatgcacctgcaccaacttcgagaaatg	1545
Db	1483	GAAAAATTTTGTGAATAAGAAATTTAAATTAATTAATTTGACAAAGATCGTGATGATTAATA	1542
QY	1546	atgggaagaacggtactcttctaacaagfvgggagaactgcacacataataacttcctgtaccc	1605
Db	1543	TTCACTGCAGAGATTAACATTTAATGTTTGAAAAACAAAAATATAATTAATTAATTTTCATCC	1602
QY	1606	falgagaattctaaacataactcttvgaaagctccaccaagctcttiaagatalatvgagac	1665
Db	1603	TCTATATATTCTGTATFATTAATGTTCAAAAAATTAATAAAGCTCTTTCATATCTTGAAAT	1662
QY	1666	tattctcgcggaacatgctgttgggagaagaactaagatttaacaagaatctctaagt	1725
Db	1663	TATCTCTTAACAAAAAGAAATTTCTGAAAAAGATTTTAATCATATTAATTAATCTTTGAANA	1722
QY	1726	aagatlcgaaacagagatcgcagacgctgttctggaacattaaagaagatfagaagaacgctg	1785
Db	1723	GGCCTCGAAGCTGATTAATAAAAAATTAAACGAAAGAAATTAAGAGTAGTGAAAAACAAAT	1782
QY	1786	tttggagaagaatttacaagaagcgaataataaccgatfagaagaatccttggaggtcc	1845
Db	1783	TTAGAAAAAATTTTAAAGACTTAACACATTCGCAAAAG---CTTCTTAACAAATTA	1839
QY	1846	gatatgtttaaagtcagcagcagagctgtcctcctatgaacagatfagtgaactcaag	1905
Db	1840	GATATTTGTAATAATTACAAAGTACAAAGAAAGTTTATTAATTAATAAATAATGAGACCTTAAGA	1899
QY	1906	aagactcaactatctctgaagaagctgagfitaacaacataatacatgtgcccgaatagt	1965
Db	1900	AAGATAGATATTATTTTAAAAAATGCAACAATCAAAAGATAGATTACATGTCACAAATAT	1959
QY	1966	tataagcagsgaataagcaggaacatactactactatgctactgaagaagaagfagac	2025
Db	1960	TATTAACCCACAAATTAACACAGAACCATATTAATTAATGTTTAATAAAAAAGACTAAT	2019
QY	2026	aaactgaaagfgttcatgcccacaagtcogagagcttcatcaacgaagaagaaagaacatt	2085
Db	2020	AAATTAATAAGAAATTTATACCAAAAGTAAAGACATGTTAAA--GAAGAACAAAGCTGCT	2077
QY	2086	aaaactgagaagcagfcaagataactccgagcttccacaagaagsgagataacccgagaag	2145
Db	2078	TATCAAGTATTACCAACCTTTAGTTGGACCAAGCAACCAACTGAGAT--GGGGGTAC	2136
QY	2146	gtctaccaccaagcccggaacaacagccggtctcagctctcgaagcgatagcgtgaagct	2205
Db	2137	TCCACACACACATTAATCCCAATCAGGAGAAACGAAGAACAAGAAACAAACACTAGACAC	2196
QY	2206	caagcagaagcagaagcagcagacagcctccagfgcagfgcagfgcocggttccagaagctaa	2265
Db	2197	ACAACAAAGGTAATCAATTAACATTACACACCAACCACTTCACCAACCAAAAGAGTAA	2256
QY	2266	gctcaagfgccttaacacacagctcctctgtaataaacaagaccggaatgctgaacaaactg	2325
Db	2257	GTTGTTGAATAATTCATTAATGAACATTAAGAGTAAATGACATTAATTCACAAAGCTTGACAAAA	2316
QY	2326	gactaaccttggaaagcctttagcttcccttaataacttcaatcatctgcacaacataatc	2385

[illegible]

D	3394	GAGATATCAATTCAAACAGAAAGATAATTATGCCAATTTAGAAAAATTTAGACTATTAAAGT	34353
Q	3388	aagctcgaaagcgaagctgtaagagaacaactgaaacctgtagaagaagaagctcgaactcc	34473
D	3454	AAAAATAGATGGAAAAACCTAAATGATAATTTTCATTTTGGAAAAAATAATATCTTTCTTA	35133
Q	3448	tctagcgactgcgtctaccctgcgctgcgcgagcttaagaagtcatttaagaacaagaactac	35073
D	3514	TCAAATGGATTACATCACTTTTAATTTACTGAAATTTAAAAAGAACTAATATAAAATATATAT	35733
Q	3508	accgcgaatagcccaagcagaaataatacagaacgtaataaagcaactgtaactcttacaag	35673
D	3574	ACAGGTATTTCTCCAACTGAAATAAATAAAGAAAGTTAACGAAGCTTTAAATCTTACGA	36333
Q	3568	aagtcctgcctgaagaaacaagatgctgcacatgctgctgc-----	36083
D	3634	AATTTTCTCCAGAAAGCAAAAGTTTACAACTGGTTTACTCCACCTCAACACAGATGTAAT	36933
Q	3609	-----tgaatcgtctccgacacacactgcggagcgtctcaactcaagaa-----	36503
D	3694	CCATCTCCATTATCTGTAAAGGTAAGTGGTAGTTCCAGGATCCACAAAGAAAGAACACAA	37533
Q	3651	--gcctgcatctactcatgctgcgagcccgagctcaatatacatcaacactcgaagctc	37083
D	3754	ATACCAATCTCAGAGCTCTTTATTAAACGAATTTCAACAAGTGTACAAATCAACAAAAATAT	38133
Q	3709	gacgatagagctcgatgacgcgtcatctgctcattcttcggagagagaaggaaggaactac	37683
D	3814	GACGGAAGAGTGTGATCTCTTAGTTGTATTAACCATTTTGGAGAAATCCGAAATATATGAC	38733
Q	3769	gatgaccccgccagcagctgctcaccgctgagagcgtcactccctccgtgatitagaact	38283
D	3874	GAAATTTAGATCAAGTAAGTAAGTGGAGAAAGC---AATATCTGTCAATGATATAATATC	39303
Q	3829	ctgtccaaatccggaacggaatacgaagctctatctcgtaaacctctgycagcgctcat	38883
D	3931	CTCTCAGGATTTGAAATATGAAATATGATGTTATATTTTAAACCTTTACCTGGACTATAT	39903
Q	3889	aggtctctcaagaacacgcctggaagaataacgtatagactcctaagtcaagctgaagagc	39483
D	3991	AGAGCTTAAAAAACAATAATTTGAAAAAACAATTTTACATTTAATTTAAATTTGAGCAT	40503
Q	3949	attctgaacagcgccttaataagaagagaaattcaagaacgctctgtagagcgacttg	40083
D	4051	ATCTTAATTTCACTCTCTTAAAGAAACGAAATATTTCTTGATGATGTATGTGAATCTGATTTA	41103
Q	4009	attccctataaagccctgacctctctctaactagctgtgtaaggaacctatacaagctctc	40683
D	4111	ATGCATTTTAAACATATATATCTCTCAAAATGAATATATTTGAAAGATTCATTTAATTAATTG	41703
Q	4069	aataaaggaagaggaataattctgctcagcttacaactatatacaagactccatcgac	41283
D	4171	AATTCAGAACAAAAAACAACACTTTTAAAAAGTTACAAATATATATAAAGATCACTAGAA	42303
Q	4129	accgatatacaattcgcctaatgatgctgctggtgatitacaagaactcctgagagaataaac	41883
D	4231	AATGATATTAATTTTGGACAGGAAGGTATATAGTTATTAAGAAAGGTTTTACGCAAAATAT	42903
Q	4189	aagctgcacctgactctataaagaatatacaacgaataag-----	42303
D	4291	AAGGATGTTTAGAATCAATATTTAAAAAGTTATCAAAAGAAAGAAAGAGTTCCTCATATCA	43503
Q	4231	-----caagcgaagaatgaaataat	42513
D	4351	CCACCAACACACCTCCGTACACAGCAAAAAACAGAGCAAAAGCAAAAGCAAAAGTAAAGTTTC	44103
Q	4252	ctgcacctcccgtaatacatcgaaacctgttaagaacagtgtaacgaacaaatcgactc	43113
D	4411	CTTCCATTTTTTAAACAAACATTGAGACCTTATACAAATACTTGTTAGTTATATAAATTGACAT	44703
Q	4312	tctgtaattcaactcggagagcccaagctcctaactataacttaacgaagaagaacatgtgaa	43713
D	4471	TACTTAATTAACCTTAAGGCAAAAGATTAAAGATTTGTAATGTTGAAAAAGATTAACCATAT	45303

OY	2140	ggacaggctaccaccaaagcccgagacacagccggttcagcttcgaaagcgatagcgtg	2199
Db	2275	ACACACACATTTATCCCAATCAGAGAGAAACAGAGTACAGAGAAACAGAGTAAACAGAA	2334
OY	2200	caagctcaagcagacaaagcggaagcgagcgacagctccagctgcagtcgagtcgccgttcc	2255
Db	2335	GAACACAGTAGGACACACACAAACGCTAACATTAACCTTTCACACCAAAAGAAAGTAACAGCA	2394
OY	2256	-----agaggtctaaagctcaagctgcgtctacacccacccagctccctgtgaataacacgcag	2310
Db	2395	CCAAAAGAGTAAAGTGTGTGAATAATTCATATAGAACATTAAGATATAGACATTTCCAAA	2454
OY	2311	aatgtcagcaaacctgcgtactccttggaagctctatgagctccctgaatacactcactc	2370
Db	2455	GCCCTTGACAAAACAGCTTTCTTCTAAAGAAATTGATGAATTTTTCATTAATATATATA	2514
OY	2371	tgcacaaatataccctccgctctccagacactatgaagagaagatctcttaaacagctac	2430
Db	2515	TGCTATATATATATTTTATGATCAAACTCAGTATGACCAAAAATTATTAGAGGTATAT	2574
OY	2431	aagaataaccaaagagagaglaaacgtctcctctgtgacccactgcgacctcgtgtc	2490
Db	2575	AATCTTACTCCAGAAAGAAAGAAAGAAATTAATTCATGTGATGCCATTTGATTTATTTAT	2634
OY	2491	aatatccagaaacaacatcccgctgatagtatctatgtcttgatagcctcaacatctctc	2550
Db	2635	AATATTCAAAATTAACATACCTCGTATGATTCATTATATGATATGATGAACAACTATTTA	2694
OY	2551	lctcaactgtctcatgagagataatgaaagagatgtctgcacactgataaacctcaa	2610
Db	2695	CAACATCTCTTTTGTGAATTAATATCAAAAGAAATGATTTATTTATTTACATTAACCTAAA	2754
OY	2611	gacaaagcaagaltgaagacctctgagag-----	2641
Db	2755	GAGGAAATTCACATCAAAAATTTATTAAGAGACCAAAAACAATTAACGTGAACATCATCT	2814
OY	2642	-----aagctaaagaagctccacactctgttaacactctctc	2681
Db	2815	ACATCCAGTCTGGAAATTAACACCGTAAATACGTCTCAATCCGCAACTCCACGTAATTTCC	2874
OY	2682	cagctccatgtcaaacactgtctctcaaacctcaagca-----	2720
Db	2875	CAAAAACCAACATCAATTCGATCTCTACCAATACCCAAAATGGTGATGTATCATCT	2934
OY	2721	-----gccgaagtgaagcgtcaagcagcacctctcactgcgacccaacctataactca	2775
Db	2935	GGTCTCGTGTAGTTGAAGAAATGATCAATGACCTTAAACGTATTGTCTATTAATGTAACAT	2994
OY	2776	ctgaagactgttgaaacatcccgctctccggaagaataaagaacat---ctacaaaga	2832
Db	2995	TTGAAAGGTATTTGTAGTCTCTTAAATCTTTGGAAATTAACCTTAAGTACTATTCATTA	3054
OY	2833	cttatgtgacagaaatcgtccggaacctctacagaaagatactgaaagacagcgacaca	2892
Db	3055	ACCATTTCTACAAACAGATGAGAAAAATTTTATGAGAAATTTTAAAAAATATGATATCC	3114
OY	2893	ttctataagagagcttcactaactcgtgaaatctaaagccgaatgatatacactctct	2952
Db	3115	TATTTTATGTGATTAACAAACATTCGTAAATCTTAATTTCAAAAGTAAATTAACAGTTTG	3174
OY	2953	aagatgaatctaaagctaaagcctggaagagacatcaatcaagctgaagaagacgtg	3012
Db	3175	ACC---GAACACCAAAAAATTCGATTAAATGATGAATTAATAAAATTAAGATATCTTTA	3231
OY	3013	caactgaagctcgacactgtacaaacagtaacaaactgaacctggaagctcttcgacag	3072
Db	3232	CAGTATCATTTGATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG	3291
OY	3073	aagaagacagctcggaagctataagatgcagataaagaagtgtacctctgtcaagagcag	3132
Db	3292	AAAAAAGAACTTGCCCAAGACCAAAATGCAAAATTTAAAAAACTTACTTTATTAAAAAACCA	3351

QY	3133	cttgaagaacactcaactctacatcgaaacatcgaaacacgtaatctgcgaactcttcagtg	3192
Db	3352	TTAAACATCAAAATTGAATTCATTAAATTAACCCACATAAATGATTTACAAAACCTTTCTGTGTT	3411
QY	3193	tctctcaacaagaagaagcgcgaatcgccgagacagaaacactctggaacacc	3252
Db	3412	TTCTTTAAACAAAAAAGAAGCTGGAATATAGCAAACTGAAACACTTATGAAAAACGA	3471
QY	3253	aagctctctcaaacactcaaaagccctgcgaatlatlaatgtagcagctctctcc	3312
Db	3472	AAAATATATATGAAACCTTTTAAAGACCTGTGTTAAATTTTAAATAGGAAACATCTTCA	3531
QY	3313	ctgaagactctctccgagagagcctccagacccgagataactatagccgaagccctcgagac	3372
Db	3532	TTAAAAACTTTAAGTGAAGTATCATCAATTCAACAGAAAGATATATATGCCATTATGAAAAA	3591
QY	3373	tccaagtcctcgtctaaactcgaaagcaagctgaagacaacctgaacctggaagaag	3432
Db	3592	TTTAGAGCATTAATCAATTAATAGATAGAAAACTCAATGATAATTTACATTTAGAAAAA	3651
QY	3433	aagctcagctaccctcttagagacgactaccctacgtatcgagagctgaagaagctat	3492
Db	3652	AAATATATCTTCTTATATCAAGTGGATTACATCATTTATTTACTGAATTAAAGAACTAATA	3711
QY	3493	aagacaagaactcacacgggcaatagcccaagagagaaataatacagacgtgaatacga	3552
Db	3712	AAAATTAATAATTTATACAGGTAATTTCCAAAGGAAAAATATATAGAAAGTTAACGAACCT	3771
QY	3553	ctggaatcttcaagaagcttcgcgctcgaaagaaacagatgctccacgtgtagtc----	3608
Db	3772	TTAAAAATCTTACGAAAAATTTTTCACGAAGCAAAAGTTACAAAGTTGTAACTCCACCT	3831
QY	3609	-----tgatctgctcgcgacacacgtgagcgagctca	3641
Db	3832	CAACCAAGATGTAATCCATCTCCATTAATGTGAAGGTAAGTGGATGTTACAGATCCACA	3891
QY	3642	acctaaaga-----gcctcatctactcatgctcgagccgagccaatacaattacc	3693
Db	3892	AAAGAGAAACACAAATTAACCAACTTCACGGCTCTTTTAAACGAATTTACAAACACTAGTA	3951
QY	3694	acatctcgaagcgcgcgcgaatgagctgcagatcatatgctcctatctcgagag	3753
Db	3952	CAATTACAAAATTTATGACGAAGAAGATGATCTCTTAGTTGCTATTACCAATTTTGGAGAA	4011
QY	3754	agcgaagagactcagatgacccctcgccagtggtgtaaccgctggaagctgtaactctcc	3813
Db	4012	TCCGAGATTAATGACGAATTTTATAGATCAAGTGTACTGCGGAAGC---AATATCTGTC	4068
QY	3814	gtgatgataaacctctgctcaaacatcgaaagaaatacgaagtgctcatctgaacct	3873
Db	4069	ACATAGATTAATATCCCTCCTCAGATTTTGAATAAGAAATATGATGTTATATTTAAACCT	4128
QY	3874	ctcgagagcgcgtcataaggtctctcaagaacacgctggaagaataacgtgtagccctcat	3933
Db	4129	TTAGCTGGAGTATATTAACAACCTTAAAAAAAACAATTGAAAAAACAATTATTTACATTATAT	4188
QY	3934	gtcaacgtgaaagcactctcgaaacgcgcgttaataaagagaaatcttcaagaagctc	3993
Db	4189	TTAAATTTGAACGATATCTTAAATTTCAACGCTTTTAAACGAAGAAATATTTCTTACATGTA	4248
QY	3994	cttgagagcgcagctgattctccctlaaagacactgcctctcaactacgtctgcaagac	4053
Db	4249	TTAGATCTGATTTAATGCAATTTAAACATATATCTCCAAATGAATACCTTATTTGACAT	4308
QY	4054	ccatacaagctccctcaataaagaagagggabaaattctgcttagtacaatatac	4113
Db	4309	TCATTTAAATTAATGAAATTCAGAAACAAAAAACACACTTTTAAAAAGTTACAAATATATA	4368
QY	4114	aagagctccatcgacaccggtatcaatttcgctaagatgctgtagtggtatataaagtc	4173
Db	4369	AAAGAAATCAGTAGAAATATTAATTAATTTTGCAACGAAGAGGTATATAGTTATATGAAAG	4428
QY	4174	ctgagcaaaaatcaagctctgagccctgtaactcattaaaaatatacaagataag---	4230

Db	4429	TTTTAGGGAATATATAGGATGATTAGAAATCAATTTAAAAAGTTATCAAGAGAAAG	4488
Oy	4231	-----ca	4233
Db	4489	GAGAGTTCCATCATCACCAACACCTCCGTACACGCAAAAAACAGACGACAA	4548
Oy	4234	ggcggaaatgaaaaatctgcgccttcctgaaataacatcgaaacctgtacaagaatg	4293
Db	4519	AAGAAGGAAAGTAAATTCCTTCATTTTAAACAAAATTGAGACCTTATACATAACTTA	4608
Oy	4294	aacgcaaatctgcacctctgttaattcaaccggggcccaaggtctcaactacttac	4353
Db	4609	GTTAATTAATATGACGATTACTTAATTAACTTAAGGCAAAATTAACCATTTGTAATGTT	4668
Oy	4354	gagaaagcaatgtlgaagltlaaaaatcaagagcgtgaacttacccaanaalccaaag	4413
Db	4669	GAAAAAGATGAAGCAGATGTTAAATTAATCAAACTTAGATTAATTAAGCAATTTGATGAC	4728
Oy	4414	aagctgcagatctcaagaaaaatacaaatctgcgcgaattgcgaactgtctacgat	4473
Db	4729	AAAAATGATCTTTTAAAAACCTAAACGCTTCGAAAGCAATTTAAAAATTTGATTAATGAT	4788
Oy	4474	tataaccacaacatctctcgtgaccaagttctgctcaactgcgagatgltgtcgaaaacc	4533
Db	4789	GATACGAAAAAAGATATGCTTTGGCAAAATTAAGTTACTTACGAGATTAG--TTCAAAATTTT	4845
Oy	4534	gccaaacagtgctgaagcaatctgctgcgagcaacctgcgaagagatctgacatctcc	4593
Db	4846	CCTAATACAAATTAATCAAAATTAATTAATGAGAAATTCGAAGATATGTTAAACATTCA	4905
Oy	4534	cagcacaatgcgtgaagaaaaagatgcccccaagaatagcgcgtgttcaagcatctgac	4653
Db	4906	CAACCCAAATGGGTAAAAAACAATGCTCGAGAAATTTCTGATGTTTCAGACATTTAGAT	4965
Oy	4654	gagcgcgaagatgtaagtgtctcctgacatacaaaagaagagatataagtcgtgag	4713
Db	4966	GAAAAGAAAGATGTAATGTTATTAAATTAACAAACGAAGGATGATTAATGTTGTA	5025
Oy	4714	aaccacaacctactcgaatgaaatacaatgctgggtgtgacgcgatctaaatgacac	4773
Db	5026	AATCCAAATCTACTGTTGATACGAAAAATTAATGTTGATGATGATGACGATGCCACATGTACC	5085
Oy	4774	gaggaagacagcgcctctcaacggaanaagaanaalacacatcgagtgtaactaagccgactcc	4833
Db	5086	GAAAGAAAGATTCAGTAGACGACGAAGAAAGAAATCACATGGAATGATTAACCTGATCT	5145
Oy	4834	tatcaactctcgcgcggaatttttgcctccagctcaaatcttcctcgtgcaatctcctctgt	4893
Db	5146	TATCCACTTTTCGATGGATTTTCGACAGTTCCTTAACCTTATAGGAATATCATCTCTTA	5205
Oy	4894	ctgacccatcgtgctgactcgtgaagcatctcaatgaatgatgag	4939
Db	5206	TTAATACATGTTAATTAATTAACAGTTTCATTTAAAAAATGTAGG	5251

RESULT 13

LOCUS DEFINITION	PFAM565A	5754 bp	mRNA	linear	INV 14-MAR-1994
ACCESSION	P.falciparum major merozoite surface antigen (PFM5A)			mRNA, complete cds, isolate FC27.	
VERSION	M9143				
KEYWORDS	M9143.1	GI:160412			
SOURCE	antigen: glycoprotein.				
ORGANISM	P.falciparum (isolate FC27 from Papua New Guinea), clone Ag75, g1.1, g126, pepG3.3.				
REFERENCE	Plasmodium falciparum				
AUTHORS	Eukaryota: Alveolata: Apicomplexa: Haemosporida: Plasmodium.				
TITLE	1 (bases 1 to 5754)				
	Peterson,M.G., Coppel,R.L., McIntyre,P., Langford,C.J., Woodrow,G., Brown,G.V., Anders,R.F. and Kemp,D.J.				
	Variation in the precursor to the major merozoite surface antigen of Plasmodium falciparum				

[illegible]

QY 190 gccgttacacaacgaacacccggctctaaagggtctgtggttagcggtgtgctccggtg 249
DB 301 GCGTTACACACTAGTACCTGGTTCAGAGGTTCACTTACTTACAGTGGTTCAGTTGCT 360
QY 250 tctgtgctcttggtgggtctgcgcctccggcgagcggtgtgcatcaggtgtgctga 309
DB 361 TCGAGTGTCTCACTTCTTC-----AGGTGTTGAGGTGGCTCAGTT 402
QY 310 gaaagcggtgtccgggaacagtcgaagaacacatccatctgtacacatctgaatcc 369
DB 403 GCTTCAGGTGGTTCAGGTATATCAAGACGTAACAATCTTCAGATTAATTCAGTATCA 462
QY 370 gacgcgaagctcgaagcgcgaacctcaagcagctgtgaagaactatctctcaatca 429
DB 463 AATACATAAATACCTGATTTAAACATAGAGTTCACAAATTAATCTGTTCACTATTAA 522
QY 430 gaagctgaatcaccacagttgttcgaacctcaatcatatgtctgacactgtgtga 489
DB 523 GAACCTAAATATCCGAACCTTTGATTTAACCAATCATATGTGACTTTATTTAAAAAT 582
QY 490 attcatgtctcaatatctgtatgtacggttcaagaagatcaatgaactcctgtaca 549
DB 583 GTTGATGGTTTCAATATTTAATGATGATGATGATGATGATGATGATGATGATGAT 642
QY 550 ttgaattctactcgaactgtcgaaggccaaactgaatgaagttgtgcgaactgact 609
DB 643 TTAACCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 702
QY 610 tctcaatctccatctcaatttgaagatcagagccagagctgtgagctatgtgaaga 669
DB 703 TGTCAAAATACCTTTCAATCTTTAAATTCGTGCAATGCAATTTAGACGTACTTTAAA 762
QY 670 gtcttcgatatcgaagcctctcgaacaacatcaagaagacatgttgggaagatgtga 729
DB 763 GGTTCGATATGTGAAGAACCATTTAGACAAATTTAAAGATATGTAGGAAAAATGGA 822
QY 730 tatataaagaataagaagaccatcgaagacataaagaagctgtgtcgaagaatcc 789
DB 823 TACATTTAAATAATTAACACCATAGCAAAATTAATGAATTAATTTGAGGAGACTTAC 882
QY 790 aagaccatagacaaaataagaatgcaacaaaggaagaagaagaagatgtgtacc 849
DB 883 AAACATTTGATCAAAATTAAGATGCAATATGAGAGGAGGAAAAAAATTTATTC 942
QY 850 gccagctacgacctgtcatctataaacaacagcttgaagaagccataacccatcag 909
DB 943 GCTCAATATTAATCTTTTATTTTACAAATTAACAAATTAACAAATTAATTTAAT 1002
QY 910 gtaagtgagaagcgcatagacaccccaagaagaatgaagaatcaagaagctgtcga 969
DB 1003 GTTTTGAATAACGTATGTGACCTTTTAAATAAATAAATAAATAAATAAATTTACT 1062
QY 970 aagattatgaataatagaatctctcgcga---gccaaactctgtgaagacccctaa 1026
DB 1063 GATATGATTAATAATTAACAGATGCGGAAACCCCACTACTGAGATTAACCAATTCCT 1122
QY 1027 ctgtgtgacaagaacaagaatagagag---cacgagaagaatcaagaagctgcgc 1083
DB 1123 CTCCTCGAATTAAGAAAAAGAGTCGAGGACAGAGAAAAATAAAGAAATTTGCC 1182
QY 1084 aaacacataatgaatcaatagatctctcttcaatgactcccttgagctgtgactac 1143
DB 1183 AAAACATTAATTAATTAACATTTGATTTTATTTATTTATTTATTTATTTATTTAT 1242
QY 1144 ttgagagagaagataagaatagatcctcgcgaagctga-----gacaag 1194
DB 1243 TTAAGGAAAAAATTAATAAGTTGATTAACCTTAATTAACCAATCTCTAGAAATCT 1302
QY 1195 gaatcagcaactaataatgaatcctcaatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1254
DB 1303 GTTCAAAATACCAAAAGTCTCTTATCCAAATGATATGATATCTTTACACACTGAT 1362
QY 1255 atcaacaacgctctc-----aaagagctcaatagcttggtgacttgatatacc 1305

DB 1363 ATTCATTAATTCATTAACCTCAGATAATGATTAATAATTCATATGCTATTAATGAAT 1422
QY 1306 ttcatatatacgaagaacccctcaagaatctacacagacaaatgaagaagaagttc 1365
DB 1423 GATACATAAGAAAAATTAATGAATAATTTATTCAGATTAATTAAGAAAAAATATTC 1482
QY 1366 atcaacgaataatcagaagaagaatcaaatctgaagaagaanaatltgaggtgaaga 1425
DB 1483 ATTAATTAACATTTAAATAAATAATTTGATTTAGAGAAAAAAACATTAATCAACAA 1542
QY 1426 agttcgaagaccgcgcgaagaagctcaacgatatcactcaagaagatgaagaagct 1485
DB 1543 CAAAATTAATAATTTACTGAAGATTAATGAAGCTCAAAAAGGATTAATGAAGATTT 1602
QY 1486 aacgagatctatgtatccaatcaacaaataacatgcagctgtcgaactctgaaga 1545
DB 1603 GAAAAATTTTATTAATAATGAATTTAATTAATTTTACAAAGATCTCTAGATTAATA 1662
QY 1546 atgggaacacgylactcttacaagaatgtgaacatgtcacacacataatccttgc 1605
DB 1663 TTCAGTCAAGATTAATCATATATTTGAAAAAACAAGATTAATTAATTTTCATCT 1722
QY 1606 tatgaatattcaagataatcttgaagaagctcaccagaagctcttaagtatatgag 1665
DB 1723 TCTAATTAATTCGTATATATGTTCAAAATTTAAAGGCTCTTTCATATCTTGAAG 1782
QY 1666 taticctgcggaacatctgtgtgaagaagaactaaatlatatacgaatcctcaat 1725
DB 1783 TATTTCTTAAGAAAGAAAGATTTCTGAAAAAGATTTTAATCATTTATTTACTTTGA 1842
QY 1726 aagatcgaacacgagatcgaagcgtctgtgagaacataagaagaatgaagaagct 1785
DB 1843 GGCTCTGAAGCTGATTAATAAATAATTAACGAAGAAATTAAGGTGTGTAATAACAA 1902
QY 1786 ttgagaagaagaatctacaaagaacgaataaaccagatgtgaagaatcctcgtgag 1845
DB 1903 CTGAAAAAATTTTAAAGGACTTAACACATTCAGCAATG---CTTCCTTAGAAGAT 1959
QY 1846 gatattgttaagltccaagltgcgaaggtgtcctcatatgaacaaatgtatgaact 1905
DB 1960 GATATTTGTAATTAATTAACATACCAAAAGGTTTATTAATTAATAAATAATTAAG 2019
QY 1906 aagatcccaactcttctgaagaagctgtgaggttaacataatataatgtgtgcga 1965
DB 2020 AAGATGAATTAATTTTAAATAATGCAACATTAAGAGATTAATTAATTAATTA 2079
QY 1966 tataagcagagaataagcagaagaccatactactcactgtactgaagaagaatag 2025
DB 2080 TATTAACCAAAATAAACCAGAACCATATTTATTTATTTATTTATTTATTTATTTAT 2139
QY 2026 aaactgtgaagltctatgcgaagctcgaagctgtatcaacgagaagaagaacat 2085
DB 2140 AATTTAAAGAAATTTATACCAAAAGTAAAGACATGTTAAAGAAAGAACAGCTGT 2199
QY 2086 aaactgaagagcgtc-----agataactcgaagcttccacagaagaagataacc 2139
DB 2200 TCAGATTAATTAACCAACCTTTAGTGCAGCAAGGAAACCAATGAAATGGGGTCA 2259
QY 2140 ggaagcgtcacaacgaagccggaacagagccggtgtgagctctcgaagaagcgtg 2199
DB 2260 ACACACACATTTATCCCAATGAGAGAAACAGATTAACAGAAAGAAACAGAAAGAA 2319
QY 2200 caagctcaagcaacaagaagaagcagcagcgtcctcaggtccaggtgtccgttcc 2255
DB 2320 GAACAGTAGAGACACACAAACAGATTAACATTAATTTACCAACAAAGAAAGATCAG 2379
QY 2256 -----agaagctaaagctcaagltgtctacacacgaagcctcgtgtgaata 2310
DB 2380 CCMAAAGAGATTAAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2439
QY 2311 aatgtcgaacaaactgtactccttgaagaagctctatgaagtctgaataatcctc 2370

Dh 2440 GCCTTGACAAAAACAGTTTATCTAAGAAATTTAGATGAATTTTAACTAAATCATATATATA 2499
Oy 2371 tggccaaatatatcctgctctcacaagcactatgaacggaagatctcttaaacagatc 2430
Dh 2500 TGTCTTAATATATTTTATGATCAAACTCTAGTATGACAAAAATTTTGAAGGTATAT 2559
Oy 2431 aagataacgaagagagagagatlaaactgctcctcttgatcaccctgagccctgctc 2490
Dh 2560 AATCTTACTCCAGAAAGAAATTAATTAATATGATGATCCATTGATTTATTTATTT 2619
Oy 2491 aatatcgaacaaacatcccgctatgctatctatgctgaagcctcaacaaatctctc 2550
Dh 2620 AATATTCAAAATACATACCTGCTATGATCATATATGATGATGACATTCATTTA 2679
Oy 2551 tctcaactgctcagtgagatatatgagaagagatgctcgaacccctataaactcaa 2610
Dh 2660 CAACATCTCTTTTGAATTTATTCAAAAAGAAATGATTTATTTACATTAACCTAAA 2739
Oy 2611 gacaacgaagaattaaagacctctgagag----- 2641
Dh 2740 GAGGAAATTCATCAAAAAATTTATAGAGAGCAAAAAACAATTAAGTGAATCATCT 2799
Oy 2642 -----aagtaagaagctccacccctcgttaaacctctctc 2681
Dh 2800 ACATCCAGTCTCGAAATACAAACCTTAATCTGCTCAATCCGCAACTCAGATTAATCC 2859
Oy 2682 cagctccatgcaacacgctctctcaccctcaagacaa----- 2720
Dh 2860 CAAAACCAACATCAAAATGCAATCTCTACCAATACCCCAAAATGCTAGCTGTATCATCT 2919
Oy 2721 -----gcccgaagctgagcctaaacgaacacccctcaccctcaacccatlaactca 2775
Dh 2920 GGTCTGTGTAGTTGAGAAAGTATGATCCCTTAACAGATTTGCTATTTAGTAAAGAT 2979
Oy 2776 ctgaactgcttgagaacatccctgctcctcggaagaatlaagaacat---ctaccaaga 2832
Dh 2980 TTGAAGGATTTGTAGTCTCTTAATCTTGAAATTAACCTAACTCAATCAATTA 3039
Oy 2833 ctctatggaacgaatctgcgagacactctcaggaagatctcgaagaacagcgagaca 2892
Dh 3040 ACCATTTCTACACAGATGGAATAAATTTATGAAATTTTAAAAATATATATACC 3099
Oy 2893 tctctacgaagagcttcaactaactcgtgaactlaaactcaacgatacaactcct 2952
Dh 3100 TATTTTATGATGATTAATCAAAATTCGTAATCTAATTCAAAAGTATTAACAGTTTG 3159
Oy 2953 aacgatgaactaaacgtaagaagctggaagagacatcaatlaagctgaagaagacactg 3012
Dh 3160 ACC--GAACACAAAAAAATGCATTAATATGATGAATTAATAAATTAAGATCTTTA 3216
Oy 3013 caactgagcttcgacctgtlaaacaagtaacaactggaacgagagactctcgacag 3072
Dh 3217 CAGTTATCATTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 3276
Oy 3073 aagaagacagctcggaagataaagctgaagctgaagctgctcgaagagag 3132
Dh 3277 AAAAAAGAACTTGGCCAGACAAAAATGCAATTTAAAAAACTTACTTTTAAAAACACA 3336
Oy 3133 ctgaagaacaaactcaactcaactgaaacaactcgaaacgtaactgagaaactctcagtg 3192
Dh 3337 TTGATATCAAAATTTGAATTCATTAAATACCCACATATGATTTCAAAAACCTTTCTGT 3396
Oy 3193 tctctcaacgaagaagacgagatgcgcgagacaggaatacaactctggaagaac 3252
Dh 3387 TTCTTTAAACAAAAAAGAGCTGAATATAGCAAACTGAAACATTTGAAAAACACA 3456
Oy 3253 aagatctctcaaacactacaagcctcgtcaagctatlaaactgagagagctctcct 3312
Dh 3457 AAAATATATTTGAACATTTATTAAGGACTTGTTAATATTTAATATGATGATCATCTCA 3516
Oy 3313 ctgaagactctctcgcgagagagatcagacagcgagaactacagcagcctcogaac 3372
Dh 3517 TTTAAAACTTTAAGTGAAGTATCAATTCAAACAGAAATTAATTAATTAATTAAGAAAA 3576

Oy 3373 ttcaagctccctgctaaagctcgaagcgaagctgaagcaaacctgaaactgagagaag 3432
Dh 3577 TTTAGAGCATTTAAGTAAATATAGTAAATGAAACTCAATGATTAATTTACATTTAGAAAA 3636
Oy 3433 aagctcaactacctctcctcagcgagactgacacctgatacgcgcgaagctcaagaagctca 3492
Dh 3637 AAAATATCTTTCTTTATCAAGTGGATTTACATTAATTAATTAATTAATTAATTAATTA 3696
Oy 3433 aagaacaaagactacacgcgcaatagcccaagcgagaataatcacagcgtgataaagca 3552
Dh 3697 AAAAAATAAAAATTAATACAGTAAATCTCCAAAGTGAATAATTAAGAAATTAACGAAGCT 3756
Oy 3553 ctggaactctcaagaagctctcgcctgaaggaagaagctgcccacgctgctgc----- 3608
Dh 3757 TTTAAATCTTACGAAATTTCTCCCAACGCAAAAAGTTACAACTGTTATCTCCACT 3816
Oy 3609 -----tgaatctgcccgcagacacactgagcagctca 3641
Dh 3817 CAACCAAGATTAATCTCCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 3876
Oy 3642 accctaaga-----gcctgcatctactcatgctcgagccgagctccaataaactacc 3693
Dh 3877 AAAGAAAGAAACAAATACCAACTTCAGGCTCTTTATTAACAGAAATTAACAAAGTACTA 3936
Oy 3694 acatctcgaagcgtcagagatgaggtcgatgacgcatcatgctgctatctcgcgag 3753
Dh 3937 CAATTAACAAATTAATGAGAAAGATATCTTATGTTATGTTATGTTATGTTATGTTATG 3996
Oy 3754 agcgaagagactacgaagacctcgccagctgctcagcggctgaagcctgacactcctcc 3813
Dh 3997 TCCGAAGATTAATGAGATTAATTTAATGATCAAGTAACTGGGAAGC---AATATCTGTC 4053
Oy 3814 gtgatgataacatctgctccaataatcgaagaagaaatcgaagtgctctatcgaacct 3873
Dh 4054 ACAATGATTAATATCTCTCAGAGATTTGAAATGAAATGATGATTTATTTAAACCT 4113
Oy 3874 ctggaagcgtctatagctctcctcaagaagaagcctggaagaatcagtgatgacactcat 3933
Dh 4114 TTAGCTGAGTATTAATGAAGCTTAAAAAACAATTTGAAAAAACAATTTATTTAAT 4173
Oy 3934 gtcaacgtgaagagacatctgaacagccgcttlaaataagaagaataatctcaagaagctc 3993
Dh 4174 TTAATTTGAACGATATCTTAAATTCACGCTTTAGAAACGAAATATTTCTTACATGTA 4233
Oy 3994 ttggaagagacttgatctccctataagaacctgacactccttaactaactgctgaagac 4053
Dh 4234 TTGATATCTGATTTAATGCAATTTAAACATATATCTCAAAATGAAATGATTAATGAAAT 4293
Oy 4054 ccatcaagctccctcaataaagaagaagagataaactctgctagctlaaactatc 4113
Dh 4294 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4353
Oy 4114 aagacatccatcgacacagcagatacaactctgcctaatagtctgctggtgctatcaagaactc 4173
Dh 4354 AAAGATATCAGTAAAGATTAATTAATTTGCACAGAAAGGATTAATTAATTAATTAATTA 4413
Oy 4174 ctgagcgaataatataagctgacactgacactctatlaaagaagatatacaagataag--- 4230
Dh 4414 GTTTTAGCGAAATTAATTAATGATTTAAGATTAATTAATTAATTAATTAATTAATTAAT 4473
Oy 4231 -----caa 4233
Dh 4474 GAGAAATTCCTCATCTCAACCAACAAACACCTCCGTTACCAAGCAAAAAACAGCAACAA 4533
Oy 4234 ggcgagaa tgaanaatatctgcccctctcgtlaaataacatcgaacccctgtaagaagctg 4293
Dh 4534 AAGAAAGAAAGTAAAGTCTCTTCATTTTAAACAAACATTTGAGACCTTATACAAATTA 4593
Oy 4294 aagacaanaatcgacactctcgtlaatcaactcgtgagcgaagctgctcaactaactaac 4353
Dh 4594 GTTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4653

QY	4354	gagaagagcaatgltggaaglttaaatcaagagcgtgaactacccaataccaagac	4413
DB	4654	GAAGAAAGATGACACATGTTAAATACTAATAGTATTTAAAGCAATTGATGAC	4713
QY	4414	aagctgcagatctcaagaaaataaacaatttgcctggaattgcagaacgctctacgat	4473
DB	4714	AAATAGATCTTTTAAACACTAACGACTTGAAAGCAATTAATAATGAT	4773
QY	4474	tataacacaacaatctcctgaacagttctgtccactgagctgggtgtcgaacac	4533
DB	4774	GATACGAAAAAGATATGCTTGGCAATTTACTTAGTACAGATTAG--TTCAAAATTTT	4830
QY	4534	gccaaacagctgtcgtgaacatctgcctgcagcggaacctgcagggcagctgcatctcc	4593
DB	4831	CCATATCATATATATCAAAATTAATTGAAGAAATTCACATATGTTAAACATTTC	4890
QY	4594	cagcacaatgctggaagaacagtgccccaagaatgagcgctgttcaagcactgagac	4653
DB	4891	CACACCATGCTTAAAAAACAATGCTCCAGAAATTTCTGATGTTTCAGACATTTAGAT	4950
QY	4654	gagcgcgaagtgcaagtgctcctcgaactacaacaagaagagataagtgctgag	4713
DB	4951	GAAGAGAAAGATGTAATGTTTATTAAATTACAAACAGAGAGTATTAATGTTGAA	5010
QY	4714	aaccacaacctacccgtgaatgaagaacatggcggtgtgacgacgactgaatgac	4773
DB	5011	AATCCAAATCTCTGTTAAGAAATTAATGATGATGATGATGATGATGATGATGATG	5070
QY	4774	gaggaagacagcgctcctcaagaaagaatacactgagctgactgaacccgactcc	4833
DB	5071	GAAGAAAGATTCAGGTAGACAGCAAGAAATATCATGTAATGTAATCACTGATTC	5130
QY	4834	tatcacctcttcgagggatcttctcctgaactcaattctcctgagcatctcctctg	4893
DB	5131	TATCCACTTTCATGATGATTTCTGCGAGTCCCTCAACTCTTACGAAATCATCTT	5190
QY	4894	ctgactcctatgctgctcctgtaacatctcaataagctgagt	4939
DB	5191	TTAATCTCATGTTAATATATATACAGTTTCAATTTAAATAATGATG	5236
RESULT	14		
LOCUS	PFMEZSA1B	5312 bp	DNA linear INV 10-MAR-2001
DEFINITION	P.falciparum gp190 (MSA1, MSP1, PMMSA) for precursor of major		
ACCESSION	235327	U09209	
VERSION	235327.1	GI:29795	
KEYWORDS	gp190 gene; merozoite surface antigen; MSA1 gene; MSP1 gene; PMMSA		
SOURCE	ORGANISM		
REFERENCE	1 (bases 1 to 5312)		
AUTHORS	Pan, W., Tolle, R. and Bujard, H.		
TITLE	A direct and rapid sequencing strategy for the Plasmodium		
JOURNAL	Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)		
MEDLINE	96133395		
REFERENCE	2 (bases 1 to 5312)		
AUTHORS	Tolle, R., Bujard, H. and Cooper, J.A.		
TITLE	Plasmodium falciparum: variations within the C-terminal region of		
JOURNAL	Exp. Parasitol. 81 (1), 47-54 (1995)		
MEDLINE	95354793		
REFERENCE	3 (bases 1 to 5312)		
AUTHORS	Tolle, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-JUL-1994) Ralf Tolle, Abt. Prof. H. Bujard, Zentrum		
COMMENT	fur Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer		
FEATURES	On Aug 5, 1995 this sequence version replaced gi:535249.		

source	1. .5312	
/organism="Plasmodium falciparum"		
/isolate="NF54"		
/db_xref="taxon:5833"		
/chromosome="9"		
/clone="3D7A"		
/dev_stage="blood stage"		
109. .165		
/gene="gp190, MSA1, MSP1, PMMSA"		
109. .5271		
/gene="gp190, MSA1, MSP1, PMMSA"		
109. .5271		
/gene="gp190, MSA1, MSP1, PMMSA"		
/codon_start=1		
/product="precursor of the major merozoite surface		
antigens"		
/protein_id="CAA84556.1"		
/db_xref="GI:929796"		
/db_xref="SPTREMBL:Q25922"		
/translation="MKILFLCSFLPIIINTOCVHESQELVKLELEDAVLTGYS		
LFQKRMVINEEITTKGASAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGS		
SGPSGTSISRSNTLPRSNTSSGASPPADASDASKSYADLKHRYRNLFTIKELKY		
ELFEDLNHMLTLCDDIHGFKYLIDGEEINELLYKLNFYEDLLRAKLNDVCANDCOI		
PNLKIIRANEILDVLRKLYGKRPDLNDKNDYKMDYIKKKTTIANINELIEGSK		
TIQDKNNDNEGSKKLYOAOYDLSIYNKOLEBANILSVLEKRTIDPLIKNNIKKL		
DKINEIKNDPPNSGNTPTLIDKKKKIDBEHEKIKETAKTIKPNIDSLFTDPLEBY		
YLREKNNKADVTPKSDPTKSVQIPKVPNPIVPLPTDTHNSLADNDNSYGLD		
NMPHTEKINEKIITDNEKRIPIINNKKIYDLERKNINHTKONKKLLEDYKSKD		
YBELLEKFEYKNNFNDKDVYDIJESAYITVLEKORNNKFSNNSVYVNOVLK		
ALSYLEDSYLRKGISSEKNNHYTLTKGLEADIKLEIKSENNKILEKNNKGLHS		
ANGSLSESDIVLYOYOKVLLIKIEDLRKIELFNKOAKDSIHVPNIYKPKNKEPY		
YLIVLKEVDKLEKPIPKYKMDLKKKQAVLSSITQPIVAASFTEDGSHPTTSOSG		
EVEVTEETEETEYGHATTVTITLPTDPSPEKRYENSTEIKSDNSGALTKIV		
YLKIDLEFLTKSYICHKYLLVSNSSMDQKLEVYNLTPREENELSCDPLDFNIN		
NIPAMSLYDSNMNDLOHLEFELYOKEMTYLHKLKEENHKKLEBOKQITGTSST		
SGNTTVAQASATHSNSOQSNASTTONGVAVSSGPAVESHPPLVLSISND		
LKGIYSLNLGKTRVNPPLITSTTEMEFEYNIILKNDTYENDDIKOFVKSNSVIT		
GLTEPOKNAINDEIKKIKPTLOLSPDIYKFKYKIDRLPKKKEIGDOKMOIKITLL		
KIOLSKLNSLNNPHVLONFVEFFNKKKEAIAETENTLENTKILKHYKLYVYN		
GESSPKLTLSVSIQIEDNYANLEKRVSLKIDGLNDLHGLKSLSPSSGLHLI		
TELKEVTKNNATGNSPSENNKVNELKSYENFLPEAKVTVTVVPPQDPVPSLVS		
RVSQSSGSTEKEQIPTSGLTLELOQVVOULDEYEDSLVLLIFGSEDNDEYLD		
OYVTEGALISVTMDNLTSGFENEYDVIYLRPLAGVYRSKLOKEKNIIPFNILNLI		
SLKRRKYFLDYLEDIMOFKRISSNEYIIESEFKLNSSEKNTLLKSKYIKESVEN		
DIKEAOGESTIYEKVLAKKDDELSIKVYKEKKEFPSSPTTPPSPAKTDQKES		
KPLPPLTNLETLYNNLVNKIDYILNAAKINDCNEDEAHVLTITKLSDLKAIIDKI		
DLFKNPYDEAIKILINDTKKMDLKLSTGLVONFPNTIISKLIEGFQDMLNISQ		
HOCVKKOCPENSGCFRHLDERRECKCLNLYKQEGDKCYENPPTCENNGGCDADATC		
TEEDGSSSKKKTTCCTGTPDSYPLRDGIRCSSNLFISFLIILMLIYSFI"		
BASE COUNT	2328 a	703 c 705 g 1576 t
ORIGIN		
Query Match	25.8%;	Score 1272.2; DB 3; Length 5312;
Best local Similarity	55.9%;	Pred. No. 1.1e-269;
Matches 2900; Conservative	0;	Mismatches 2018; Indels 267; Gaps 15;
QY	10	atgaatacatcttctcctcgttcattctgttttttcaatcaatactcagtcgtg 69
DB	109	ATGAGATCATATTTCTTTTATGTTTATTTTATTAATACACAAATGTGTA 168
QY	70	accacgaatccatcagaagcgtgttaagaacctggaagccttggaagatgcgcctt 129
DB	169	ACACATGAAGATATCAAGAACTTGCAAAAATCTGAAGCTTTAGACATGCAATTGG 228
QY	130	accgagatcagcctgtctcgaagaagaatggtgctgaatga----- 174
DB	229	ACAGTTATATGTTTATTTCAAAAGGAAAAAATGCTTTTAATGAAGAAATTTACTACA 288
QY	175	-----ggagcagtgagcagcgccgttaacaacagc 204
DB	289	AAAGTGCAAGTGCTCAAGATGCTGCAAGTGCTCAAGTGCTCAAGTGCTCAAGTGCT 348

Db 2506 CTAAAGAAATTAGATGATTTTAACTAAATCATATATATGTCATAAATATATTTTAGTA 2565
Qy 2392 tctcacagcactatgaagagaagatctcttaaacag tacaga taaccaaagaagagag 2451
Db 2566 TCAAACTCTAGATGAGCAAAAATTTATAGAGCTATATATCTTCCAGAGAAAGAA 2625
Qy 2452 agtaacgtctcctctgtgtacacgtgagcgtgtctcaataccagaacaacatccc 2511
Db 2626 AATGAATTTAAATCATGATGATCATTTATTTATTTAATATTCATAAATAACATACCT 2685
Qy 2512 gttatgtatctatgtctgtagccccaacatctctctcactcgtctcaggtgagagata 2571
Db 2686 GCTATGATATCTTATATATGATATGATGAAACATGATTTTACMACATCTCTTTTGAATTA 2745
Qy 2572 tatgagaagagatgtgtcgcgaacctgtataacctaagaacagaacagaagaltagaac 2631
Db 2746 TATCAAAAAGGAATGATTTTATTTATTTACATTAACCTTAAAGAGAAATCATCATCAAAAA 2805
Qy 2632 ctctctggaagaaagctagaaggtctc----- 2657
Db 2806 TTATTTAGAGAGCAAAAAAATAAATACAGACATCTACATCCAGTCTCGAATAACA 2865
Qy 2658 -----cactctgttaaacctctctctccagctccatgcaaccatgtc 2702
Db 2866 ACCGTAATTAATCTGCTCAATCCGCACTCACAGTAATTCGCAAAACGACACATCAATAATGCA 2925
Qy 2703 tctcacacctcaagaca-----gccgaagtgaagcgt 2736
Db 2926 TCCTCTACCAATATCCCAAAATGCTGATGCTATCATCTGCTCTGCTGTAGTTGAAGAA 2985
Qy 2737 aacgacgacacctctcactcgcgaaccttaactaactgaactgtgtgaagacatc 2796
Db 2986 AGTCATGATCCCTTAAACAGTATGCTATTTAGTACGATTTGAAAGTATTTGATCTCTC 3045
Qy 2797 ctgtctctggaagaaagataagaaatctc-----ctaccaagaacttaatggaagaaatcgtcc 2853
Db 3046 TTTAATCTTGGAATTTAAACTTAAGTACTTAATCCATTTACATCCCTTTCTACACAGAGATG 3105
Qy 2854 gagaactctcagagaagatactgaagaacagcagacacatctatacagagagctcact 2913
Db 3106 GAAAAATTTTATGAGAAATTTTAAAAAATAATGATACCTATTTTATGATGATGATGATGAT 3165
Qy 2914 aactctgtgaatctaaagccgagtatacaactctcttaacgaatgaatctaaagtaag 2973
Db 3166 CAATTCGTAAATCTAATTCAAAAAGTAATTAACGGTTTGAC-----GAAACACAAAAAAT 3222
Qy 2974 aagctggaagagacataaagctgaagaagacacgtgaagctcgaacgtgtac 3033
Db 3223 GCATTTAAATGATGAAATTTAAAAAATTAAGATCTTTACAGTTATCATTTTGAATTTATAT 3282
Qy 3034 aacaagtaacaaactgaactggaagacgtctctcgaacaagaagaacagtcggaagat 3093
Db 3283 AATTAATTTAATTAATTAATTAATGATGATTTAATTAATTAATTAATTAATTAATTAATTAAT 3342
Qy 3094 aagaatgcagatacaagaagtgtgactcgtccaagagcagctgtgaagaacaaactcaactca 3153
Db 3343 AAAATGCAAAATTTAAAAAATCTTATTTAAAAAAGAACAAATTAAGAAATTAATTAATTAAT 3402
Qy 3154 ctgaacaatctcgaacaagctgacgtgaagactctcagtgctctcgaagaagaagaagaa 3213
Db 3403 CTTAATTAACCCACATTAATGATTTACAAAACTTTCTGTTTCTTTTAAACAAAAAAGAAAGAA 3462
Qy 3214 gccgagatccgcagagacagagaacactctggaagaacacagaatctctctcaacaactac 3273
Db 3463 GCTGAATATAGCAAAATCTGAAAAACCATTTAGAAAAAATAATTTATTTGAACAACTTAT 3522
Qy 3274 aaagcctcgttcaatataatgagcaggtctctcctctgaagactctctcgcgagag 3333
Db 3523 AAAGGACTTGTTAAATTAATTAATGATGATTCATCTCAATTAATAAACTTTAAAGTGAAGTA 3582
Qy 3334 agcatccagacagagatacactacagcctcgaagaaactcgaagctcgtctgaagctc 3393
Db 3583 TCATTTCAACACAGAGATTAATTAATGCCCAATTTAGAAAAATTTTAGAGTATTTAAGTAAATA 3642

Qy 3394 gaagcgaagctgaaggacaaacctgaacctggaagaagaagctcagctcactctcagc 3453
Db 3643 GATGAAAAACTCAATGATATATTATTTAGAGAAAGAAAAAATATCTTTCTTATCAAGT 3702
Qy 3454 ggaatgcatacactgtatgcgcgagctcaaggaatcatataagaacaagaactacacgcgc 3513
Db 3703 GGATTTACATCATTTAATTTCTGTAATTTAAAAAGAGTAAATTAATTAATTAATTAATTAAT 3762
Qy 3514 aatagcccaagcagagatatatacagacgtgaataagcgaactgaactcttacaagaatctc 3573
Db 3763 AATTTCTCCAAGTCAAAATATATAGAAAGTTTAAGAAAGCTTTTAAATCTTACGAAATTTT 3822
Qy 3574 ctgcctgaagaaacagatgtcgcgaactgtggtgtc----- 3608
Db 3823 CTCGCCAAGAGCAAAAGTTTCAACAGTTTGAATCTCACCTCAACCATGATGTAACCTCATCT 3882
Qy 3609 -----tgaatctgctccgacacactggaagcagtlcacaactaagaag-----cct 3654
Db 3883 CCATTTATCTGTAAAGGCTAGTGTAGTTCAAGATCCACAAAAAGAACACAAATATACA 3942
Qy 3655 gcatctacatcattgtcgagccgagctcacaatactacacactcagaacgtcgcagat 3714
Db 3943 ACTTCAGGCTCTTTATTAACAGATTTACACAAAGTACTCAATTTACAAATTTATGACGAA 4002
Qy 3715 gaatcgaatgaacgtatcattgtcctatctcgcggaagcgaagagactcagatgac 3774
Db 4003 GAAGATGATGATCTTGTAGTTGATTAATCAATTTTGGAAATCCGAAAGATTAATGACGAATAT 4062
Qy 3775 ctgcgcgaagtggtgtcagcgtgtgagctgcactcctctcgtgtatgataactctgtcc 3834
Db 4063 TTATGATCAAGTATTAATCTGAGAGAC-----AATATCTGTCACAAATGATTAATTAATCTCTCA 4119
Qy 3835 aaatcgaagaacgaatacgaagtgctctatctgaaacctctgcaagcgtctcactaagctc 3894
Db 4120 GGAATTTGAAATTAATATGATGATTAATTTAAACCTTTAGCTGAGATATTAAGAAC 4179
Qy 3895 ctcaagaacagctggaagaatacgtgtgactcctaactgaacgtcgaagaacatctcg 3954
Db 4180 TTTAAAAAACAATTTGAAAAAACAATTTTACATTTTAATTTAAATTTGAACGATATCTTA 4239
Qy 3955 aaagccgcttaataagaagagaatactcaagaacgctctcgtggaagcgaactgtctcc 4014
Db 4240 AATTCACGCTTTAGAAACGAAATATTTCTTTAGATGATTAATTAATCTGTTTAATGCAA 4299
Qy 4015 tataaagacctgaacctcctcactaactcgtgtcgaaggaacctacataagtlcctcaataa 4074
Db 4300 TTTAAACATATATCCCAATGAATGATATTTAGAGATTTCAATTAATTAATTAATTAATTAAT 4359
Qy 4075 gagaagaaggataaattctctgtactgaatacaactatacaagaactccatcgacacgat 4134
Db 4360 GAACAAAAAACAACACTTTTAAAAAGTTTCAAAATTAATTAATTAATTAATTAATTAATTAAT 4419
Qy 4135 atcaatctcgttaataatgtgtcgtgggtatataagaatcctcgaagaataatcaactc 4194
Db 4420 AATTAATTTGACAGGAGGTATTAATTTATTAAGAAAGGTTTACGAAATTAATTAATTAAT 4479
Qy 4195 gacctgacctataataaagatatacaacgaataag-----caagcgagaaatgaataatctcg 4230
Db 4480 GATTTAGATTCATTTAAAAAAGTTTATCAAAAGAAAGAAAGAGAGATTTCCATCATCAACA 4539
Qy 4231 -----caagcgagaaatgaataatctcg 4254
Db 4540 CCACAAACACTCCGTCACACGCAAAAACAGACGAAGCAAAAGAGAAAGTAATGCTCTT 4599
Qy 4255 ccctctcctaaataatcgaacccgtgataagaacgtgaagacagaacatcgaccccttc 4314
Db 4600 CCATTTTAAACAACATTAAGACCTTATTAACAAATTAATTAATTAATTAATTAATTAATTAAT 4659
Qy 4315 gtaatccactggaagcgaagtcctcaactaactactaagagaagcgaatggaagatc 4374
Db 4660 TTTATTTAATTAAGCAAAAGATTAACGATTTGTAATGTTGAAATAAAGATGAAGCATGTT 4719

[illegible]

BASE COUNT	2325 a	686 c	711 g	1554 t
ORIGIN				
Query Match	25.7%	Score 1267.2;	DB 3;	Length 5276;
Best Local Similarity	56.0%;	Pred. No. 1.4e-266;		
Matches 2917; Conservative	0;	Mismatches 1998;	Indels 291;	Gaps 16;
Oy	10	atgaaatcattttctctctcgttcattcgtttttatcatataactcaagtcgctg	69	
Db	1	atgaaatcatttctttttatgctcatttctttttttatatttataacacaaagtga	60	
Oy	70	accacgaatccatcagagagcttgcgtaagaacgcgagacgtttgaaagatccgtc	129	
Db	61	acacattgaaagttatcagaagaacttgcacaaaacatgaaacgtttgaaagatccatt	120	
Oy	130	accggaatcagcgttccagaagaagaagatgctgtaagtga	174	
Db	121	acaggttatgcttttatttcattgaagaaaaaaatgacttttaaaagaaagaaattact	180	
Oy	175	-----ggagcagatgagcaagcgcttaacaaacagcacaccgct	213	
Db	181	aaagtcgaagtcacaaagtggtacaaagtggtacaaagtcgacaaagtggtacaa	240	
Oy	214	tctaaagagctctgtagcagcgttgagtcgcgttggtctgtagccttcgggggtccgc	273	
Db	241	acaaagtggtacaaagtggtacaaagtggtacaaagtggtacaaagtggtacaaag	300	
Oy	274	gcttcgcgagcagcgttgcagtcaggtgctcaatgagcagcgcggttccggaacagt	333	
Db	301	ggtacaaagtggtacaaagtggtacaaagtggtacaaagtggtacaaagtggtacaa	360	
Oy	334	ggaagaaccaatcactgcaga	355	
Db	361	acttcattcattcgttcacaaacatttacctcgttcacaaatatttcatttcgttcac	420	
Oy	356	-----acttcagagatccgagcgaagtcctcactcagcactcaagcaggaatgaga	408	
Db	421	ccagctgtagtcgaagcgaattcagatcttaattttagctgattttttaaaccacgaat	480	
Oy	409	aactctcctcactacaagaagcgttgaagtaaccacagtttgctgacctcactaatcat	468	
Db	481	aatattcgttgctcattttttaaagaactcaaatatcccgaaactcttgatttttaacaa	540	
Oy	469	atgctgacactgctgtgataacattcaatgagcttcaaatatctgattgacggttaagaag	528	

Dh 541 ATGTTAACTTGTGNGATATATTCATGCTTCAATATTTAATGATGATGAGAA 600
Qy 529 atcaatgaactcctgacaagtgaattctctactgactgctaaggccaacgaat 588
Dh 601 ATTAATGAATTTATTAATTAATTAACCTTTATTTGATTTATTAAGACCAAAATTAAT 660
Qy 589 gacgtttgcgaatgactctgcaaatctcaatctgaatctgaagtcagacaag 648
Dh 661 GATGTATGCTGATATATTTGTCAAATACCTTCAATCTTAATTTGGTGCAAATGAA 720
Qy 649 ttgagcgtatgaagaagttgcttcgcatcgcaagccttcgacaacatcaagac 708
Dh 721 TTATGAGTCTTAAATAAACTTGTGCTGATATGAAAACCATTAATTAAGAT 780
Qy 709 aatgtagaagaatggaatgatatattaagaagaagaagccatcgagaactaac 768
Dh 781 AATGTATGAAAATGGAATTTACATTAATAAAATTAACACCTTACCAATTAATAT 840
Qy 769 gagctgatacgaagatcccaaaagacatagacaanaaataagatgcaaccaagagaa 828
Dh 841 GAATTAATTTGAAGAGTAAGAAAACATTTGATCAAAATTAAGATGATGATGAAGA 900
Qy 829 gaaagaagaagttgtacagagccagtaagcctgctcatctatacaaacagcttga 888
Dh 901 GGAATAAAAAAATTAATACAGCTCAATATGATCTTTTCAATTAATTAACATTAAGA 960
Qy 889 gaagcccatcaactcagcagctgactgtagaagcgcatagacacccatcaagaatga 948
Dh 961 GAAGCATATATTTATTAAGCGTTTGAAGAAAGTATGACCTTTAAAAAATTAAGAA 1020
Qy 949 aatataaagaactgctcgacaagatlaaagaatgaatccctccgccaagcgaact 1008
Dh 1021 AACATTAAGAAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 1080
Qy 1009 ggaagacacccatcaagcgtgctgtagaagaagaagaatgagaagagaagaag 1068
Dh 1081 GGAATATACACCAATATCTCTCTGTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 1140
Qy 1069 atcaagaagatcgccaacccatlaagttcaacatagatctctctactgactgctcct 1128
Dh 1141 ATTAAGAAGAAATTTGCCAAACCTATTAATTAATTAATTAATTAATTAATTAATTA 1200
Qy 1129 gagctgagactactctgtagaagaagaatgaatatagacatcctccgccaagtcga 1187
Dh 1201 GAATTAAGATTAATTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 1260
Qy 1188 -----gacaaagaatcaaccgaactaatgaatgcccaatgigtgaagtaacct 1239
Dh 1261 GATCTACGAATCTGTTCAAAATACCAAAAGTTCTTATTCCAATGATTTGATATTCCT 1320
Qy 1240 ctgtctataacgatalcaacaa-----cgctctcaagcagctcaatagctcggt 1290
Dh 1321 TTACCACTCAGATATTCATTAATTCATTAATTCATTAATTCATTAATTCATTAAT 1380
Qy 1291 gactgattaaacctctgattatcacgaagaacctcaagaatatctcaacagaacat 1350
Dh 1381 GATTTATGAATCTGATACCTTAAGAAATTAATGAATAAATTTTACAGATTAATAG 1440
Qy 1351 gagagaagaagttatcaacgaatcaaggagaagatcaaaatlgagaagaagaat 1410
Dh 1441 GAAAGAAAATATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500
Qy 1411 gagagtgacaagaagaatlgacgaagccgacgaagaagtcataagcatcactaagaag 1470
Dh 1501 AATTCACAAAAGAAACAAAATTAATAAATTAATTAATTAATTAATTAATTAATTAAT 1560
Qy 1471 tatgaagaagctgtagacgaatctatgattccaaatccaataacatcgacctgaacc 1530
Dh 1561 TATGAAGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
Qy 1531 aacttgagaagaatgtagggaagaagctctctacaaagtgagaagcagacacccat 1590
Dh 1621 GTCTGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680

Qy 1591 aatacttgatccctatgagaatctcaagcataatctgagaagctcacaacgctct 1650
Dh 1681 AATTAATTTTCCCTCTATTAATTTCTGTATTAATTAATTAATTAATTAATTAATTAAT 1740
Qy 1651 aagatatgagaactatctctcgcggaacatgctgtgtagaagaagaactaagatcac 1710
Dh 1741 TCATATCTTGAAGATTAATTTTAAGAAAGAAATTTCTGAAAAGATTTATTAATCAT 1800
Qy 1711 aagaatctcataagtagtagcgaagaacgagatcgagacgtctgttgaagacctaagaag 1770
Dh 1801 TTACTTTGAAAACCTGGCTCGAAGCTGATTAATAAATAATTAACACAGAAATTAAGAT 1860
Qy 1771 gatgaagaagctgtgttgaagaagaatctacaagaagaagaatcaacagatgagaag 1830
Dh 1861 AGTGAAGAAATTTCTTCAAAAAATTTTAAGAGCTTAACACATTCACCAATG---CT 1917
Qy 1831 atccgtgaggtcgcgatatgttlaaagtcgaagtgtagaaggtgctcctcatgaacaag 1890
Dh 1918 TCCTTGAAGATTAATGATTTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1977
Qy 1891 attgtagaactcaagaagctcaactctcgaagaagtgtagtataacataatata 1950
Dh 1978 ATTAAGACTTAAGAAAGATGAGATTAATTTTAATAAATGACACACTTAAGATGATAT 2037
Qy 1951 catgtgcgaatgatatataagcagagaaatgaagcagaacatactactcatgctac 2010
Dh 2038 CATGTACCAATATTTATTAACCAAAATTAACCAAGAACCAATATTAATTTGATTA 2097
Qy 2011 aagaagaagatgacaacactgaagtgctcatgcccgaagtcgagaagcctgataagaa 2070
Dh 2098 AAAAAAGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2155
Qy 2071 gagaagaagaacatlaaactgaagcagcagatgaactccgagcctccagaaga 2130
Dh 2156 AAGAACAACTGCTTATTAAGATTAACACACCTTTAGTGCAGAACGAAACACAG 2215
Qy 2131 gagataccgagacagctcaaccacgaagccggaacagcggttcagctcgaagc 2190
Dh 2216 AAGAT-GGGGGTACCTCCACACACACATTAATCCCAATCGAGAAACAGAAATACAGAA 2274
Qy 2191 gataagctgcaagctcaagcagaagcagaagcagcagcagcctcagtgccagcgcc 2250
Dh 2275 GAAMCAGAAAGAAACAGAAACAGTAGACACACAAACGGTAACAATTAACATTAACA 2334
Qy 2251 gtccagaagctcaagctcaagtgctcaacacacagcctcgtgatacaagaagcag 2310
Dh 2335 CCAAAAGAGTAAAGTTGTTGAATTTCAATAGAACATTAAGATTAACCAATTCACAA 2394
Qy 2311 aatgcaagaactgtagactctgagaagctctatgagttccgtgatacatcctacac 2370
Dh 2395 GCCTGACAAAACAGTTATCTTAAGAAATTAAGATTAATTAATTAATTAATTAATTA 2454
Qy 2371 tgccacaatatatctcgtctcagcacacatcaagaagaagttcttaaacagttac 2430
Dh 2455 TGTCATTAATTAATTTTATGATCAAACTGATAGTACGACCAAAATTAATTAAGGTAT 2514
Qy 2431 aagataccaagaagaagaglaaactgctcctctgtagatccactgagcctgctc 2490
Dh 2515 AATCTTACGCCAGAAACAAAATGAATTAATTAATTAATTAATTAATTAATTAATTAAT 2574
Qy 2491 aatataccaagaacatctccgttatgattctatgtctgataagcctcaacaattctcc 2550
Dh 2575 AATATTTCAAAATTAACATTAATCTGATGATGATTAATTAATTAATTAATTAATTAAT 2634
Qy 2551 tctcaactgtcatgtagatataagagaagagatgctgcaactcgtataaactcaaa 2610
Dh 2635 CAACATCTCTTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2694
Qy 2611 gacaacgacaagaatgaagacctctcgagaagcgaagagctc----- 2657
Dh 2695 GAGGAAATTCACATCAAAAATTTATTAAGAGGACCAAAACAAATTAATCTGGAACATCATCT 2754

[illegible]

Db	3832	AAAGAAAGAAACCAAAATACCAACTCCAGGCTCTTTATTTACGAATTTACAAACAATAGTA	3891
Oy	3694	acatctcgaaacgctcgaaacgctgaagtcgaatgcatactatgctgcatactctcgagac	3753
Db	3892	CAATTACAAAATTTATGACGAAGAAGATGATCTCTTAGTTGTATTAACCAATTTTGGAGAA	3951
Oy	3754	agcgagagagactcgaatgcactcgcgcgcgaagtgctacccggtgaaagctgcactctcc	3813
Db	3952	TCCGCAAGATATATGACGAATTTATAGATCAAGTGTAGTACGGAGAAAGC--ATATCTGTC	4008
Oy	3814	gtgatgtataacatctctgtccaaaatccgaaacgaatacgaagtgctctatctgaacct	3873
Db	4009	ACAAAGATATATATCCCTCAGGATTTGAAAAAGAAATATGATATATTTAAACCT	4068
Oy	3874	ctcgagagcgctcctaaagtcctcgaagaacgctggaatlaacgtgaatgcctcaat	3933
Db	4069	TTAGCTGACGATATATACAGCTTTAAAAAAAACAATTTGAAAAAACATTTTACATTTAT	4128
Oy	3934	gtcaacgctgaagagacatctcgaaacgcgcttataaagagaataattcaagaagc	3993
Db	4129	TTAAATTTGAACGATATCTTAAATATCACGCTTAAGAAAGAAAAATTTCTTACATGTA	4188
Oy	3994	ttggagagcgagcttgatccctataaagactgcactctcctaactaogttgtcaagac	4053
Db	4189	TTAGATCTGATTTTAAAGCAATTTAAACAATATTCCTCAAAATGAATATACATTTTCAAT	4248
Oy	4054	ccatacaagttccctcaataaagagaagaggaataaattctgtcaatgatacaatac	4113
Db	4249	TCATTTAAATTTATTTGAATTTACAGAACAAAAACACACTTTTAAAAAGTTACAAATATTA	4308
Oy	4114	aagagctccatcgaaacccgaatalcaatttcgcataatgctggtgattatacaagac	4173
Db	4309	AAAGATCAGTAGAATAATGATATTTAAATTTGCACAGGAAGTATTAATTATGAAAG	4368
Oy	4174	ctgagcgaaaaatacaaatctgcacttgactctataaagaatatacaagataag---	4230
Db	4369	GTTTATGCGAAATATTAAGATGATTTTAAATCAATTTAAAAAGTATACAAAGAACAAAG	4428
Oy	4231	-----aa	4233
Db	4429	GAGAAATTCCTCATATACACACACACACACCTCCGTCACCGAAAAACAGACGACAA	4488
Oy	4234	ggcgagaatgtaaaaaatactgcgcctctcctgaataaatacgaacccgtgtacaagacgtg	4293
Db	4489	AAGAAGAAAGTATAGTCTCTTCCATTTTAAACAATATGACACCTTATACAAATAGCTTA	4548
Oy	4294	aacgacaaatcgacccctctcgtaattcacctgagagcgcaaggtccctaactactac	4353
Db	4549	GTTATATAAATTTGACGATTAATCTTAATTAATCTTAAAGCAAAATTAACGATGTATGTT	4608
Oy	4354	gagaaagacaaatgtagaagttaaaatcaagagagctgaactactctaaacaatccaagac	4413
Db	4609	GAAAAAGATGAAGCAATGTTTAAATTAACCTTAACCTTAGATTTAAAGCAATTCATGATAC	4668
Oy	4414	aagctggaagattcaagaataaatacaaatctgcctcggaattgcagacctgctacgcat	4473
Db	4669	AAATTAGTCTTTTAAAAACCAATACGACTTCGATGCAATTTAAAAAATTTGATTAATGAT	4728
Oy	4474	tataacccaacaactctccgaccgaagtctgttccactgacatggtgttcgaaacctc	4533
Db	4729	GATACGAAAAAAGATATGCTTTGGCAAAATTAATTAAGTAACAGGATTAAG---TTCAAAAATTTT	4785
Oy	4534	gccaaaacagtgctcgagcaatctgcctcgaaecgcaacctgcagggcacaatgctgaactctcc	4593
Db	4786	CCTATATACAAATATATCAAAATTAATTTGAAGGAAATTTCCAAGATATGTATTAACATTTCA	4845
Oy	4594	cagcccaaatcggtgaagaacagtgcccccagaatagggctgtgttcaggcaatcgtgac	4653
Db	4846	CACACCAATGCGTAAAAAAAACAATGCTCCAGAAAATTTCTGGAATGTTTCAGACATTTAGAT	4905
Oy	4654	gagcggaagatgcaagtgctctctgtaactcaaaacaagaagaataagtgctgtgag	4713

Dd	4906	GAAAGAGAAGATGTAAATTGTTTATTTAATTCAACAAAGAAGGGTGATTAATGCTTGGA	4965
Oy	4774	aaccacaacctactctgcaatgtgaataacaaatggcgggtgttgacgccgatgtctaagcac	4773
Dd	4966	AATCCAAATTCCTACTCTGTGAACGAAAATAATAGTGTGATGGAAGCAGATGCCAAAAATGCAC	5025
Oy	4774	gaggagaagcagcggtcttaacgcgaagaanaaacacatcgagtgactgaagccgacttc	4833
Dd	5026	GAAAGAAGTTTCAGTGTACGACAACGGAAGAANAATCACATGTGATGTACTAATCAACTTCT	5085
Oy	4834	tatccaacttcgcagcggagtllttgtccacgctcaatttccctcyggacatccctccctg	4893
Dd	5086	TATCCACTTTTCGATGAGTATTTTTCGCGATTCCTCAACTTCTPTAGCATATCATCTTTA	5145
Oy	4894	cgcagcccatgcgtgactctctgaacgcttatcacaatgatcatgacg	4939
Dd	5146	TTAATTAACCTCATGTATTAATTATACAGTTTCATTTAAAAAATGTAGG	5191
RESULT	16		
PFGE195			
LOCUS		5541 bp DNA linear INV 12-SEP-1993	
DEFINITION	P.falciiparum (Camp strain) gpi95 gene for major merozoite surface antigen.		
ACCESSION	X03831.X06887		
VERSION	X03831.1 GI:9894		
KEYWORDS	glycoprotein; surface antigen; tandem repeat.		
SOURCE	malaria parasite P. falciiparum.		
ORGANISM	Plasmodium falciiparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 3593)		
TITLE	Weber,J.L., Lehninger,W.M. and Lyon,J.A.		
JOURNAL	Variation in the gene encoding a major merozoite surface antigen of		
MEDLINE	the human malaria parasite Plasmodium falciiparum		
REFERENCE	Nucleic Acids Res. 14 (8), 3311-3323 (1986)		
AUTHORS	86205236		
TITLE	2 (bases 3581 to 5396)		
JOURNAL	Weber,J.L.		
MEDLINE	Direct Submission		
REFERENCE	Submitted (12-MAR-1987)		
AUTHORS	3 (bases 3594 to 5541)		
TITLE	Weber,J.L., Sim,B.K., Lyon,J.A. and Wolff,R.		
JOURNAL	Merozoite surface protein sequence from the Camp strain of the		
MEDLINE	human malaria parasite Plasmodium falciiparum		
REFERENCE	Nucleic Acids Res. 16 (3), 1206 (1988)		
AUTHORS	88143999		
TITLE	Data kindly reviewed (12-MAR-1987) by Weber J.L.		
JOURNAL	Location/Qualifiers		
MEDLINE	1..5541		
COMMENT	/organism="Plasmodium falciiparum"		
FEATURES	/strain="Camp (Malaysia)"		
source	/db_xref="taxon:583"		
	283..5463		
CDS	/note="pot. major merozoite surface antigen"		
	/codon_start=1		
	/protein_id="CAA27446.1"		
	/db_xref="GI:9895"		
	/db_xref="SWISS-PROT:P04934"		
	/translation="MKILFLICSLFPIINIOCTHESYQELVKKLELDEAVLTGGL LFHEKMLINEEETITYGASAQSGTSIGTSIGSYGSISGTSQAQSTSDTSQAQSGT SGSAQSGTSGTSGTSPSSRNTNTPDKVJFINIKQIDLEEKNHTDEOKNK NYLPTEIKELKPELFDLTNLMTLCDDNHGKRYLDIGEIEINLLYKLNFDDLRAK LNDVCANDYCOIPENLKIRANELVLVKLVGYRKPLDNIKDQNMEDYTAKKKTTHI ANIELITEGSKTTIDONKNMNEBGRKLYOAYDISYNQLFEAHNLIVLEKRIDI TLKNEINTELKDLINELKNPNPANSNGTPRTLDDKKKEIEBEKEIKEIAKTIKPNI DSLTPDELLEYLIERKNKVDPKSDPTKSVQIPKVPINQIVPLPLTDIHNSL LAADNKSQYGDLMNPDTREKINEKIITDNKERKIFINNKKQIDLEEKNHTDEOKNK KLDEIDYEKSKDYDEELLEKEYEEMFNENNDFDVJFSARYTNVEORYNKFSSSK NNSVYNQOLKKALESYLEDYSLRGISPKDNNHYTLTKGLEADIKRTIEIKSENNK ILEKNFGITHSNASAASEVDYIVKLOKYVLLIKKIDIRKIELFIKNAOLKSIDIVP NIYRPQNKPPTYLIYLAKREVDAKREFPKRYDKMLKQVALSSITOPVAASETTED GGHSHTLSQSGETEEETEETEETEVGHATVTITPLEPVKVENVSIEHRSDNSQG ALTTVYIAKLKIDEFLTKSYIOHKILVASNSMDRLKLEPPVLEEBENLKSCDPLDD		

[illegible]

Db 943 GATGATGCTGCTAATGATTATGTCAAATACCTTCAATCTTAAATCGTCAATGAA 1002
Qy 649 ttgagcattgagaagaagtgtgttcgtgatatcgcaagcctctcgacaacatcaagac 708
Db 1003 TTAGACGTACTTAAATAAACTTGCTGATATGATAAATCAATTAAGCAATTAATAAT 1062
Qy 709 aatctggaagaatcgagaatcatatataaagaataagaagaccatcgagaacatlaac 768
Db 1063 AATGTAGAAAAATGGAAGATTACATTAATAAAAAATTAACACCATATGCAATTAAT 1122
Qy 769 gagctgacgagaatcccaaaagccatagacaanaaiaaagatgcacacgaagagaa 828
Db 1123 GAATTAATTAAGAGAAAGTAAGAAAAATCAATGATCAAAATTAAGAAATGCAATTAAGAA 1182
Qy 829 gaagaagaagaagctgtaccaggtccagtagaactgtctcatctatacaaacagcttga 888
Db 1183 GGAATAAATAATTAATACCAAGCTCAATATGATCTTCTATTATTAATTAATTAAGAA 1242
Qy 889 gaagcccaataaccatcagcgtactgagagagcgatagacagaccctcaagaagaatga 948
Db 1243 GAAGCACAATTAATTAATTAAGCGTTTGAATAAACGTATTGACACTTTAATAAATAATGA 1302
Qy 949 aatatcaagaacgtctgcgacaagaaltgaatlaagaaacctcgccagccaactct 1008
Db 1303 AACATTAAGGAATTAATTAATTAAGGCTTTGAATAAACGTATTGACACTTTAATAAATAATGA 1362
Qy 1009 gggagaacccctcaacagcgtctgcgacaagaacaagaatagaggagcagagaagaag 1068
Db 1363 GGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1422
Qy 1069 atcaagaagatcgccaacacccatlaagtlcaacatagatctctcttactgactccct 1128
Db 1423 ATAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1482
Qy 1129 gagctgagtagtactctgagagagagaataagaatagacatctccgccaagtlcga 1187
Db 1483 GAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1542
Qy 1188 -----gacaaaggaatcaacggaactaigaatctcccaatggtgtgagctaccc 1239
Db 1543 GATCCTACGAAATCTGTTCAATACCAAAAGTCTTATTCACAAATGATTAATATCT 1602
Qy 1240 ctgctctataacgatalcaaca-----cgctcacaagcgctcaatagcttgct 1290
Db 1603 TTACACCTCACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1662
Qy 1291 gacttgatlaacccctcgaltatatacgaagaacccctlaagaatlatcacagacat 1350
Db 1663 GATTTAATGAATCTGATTAATTAAGAAAAATTAATGAATAAATTAATTAATTAATTAAT 1722
Qy 1351 gagagaaagaagttlatcaacgaatacaagagaagatcaaaatctgagaagaagaatc 1410
Db 1723 GAAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1782
Qy 1411 gagagtgacaagaagtltaacgaagccgcaagatcacaacgatalactcaagaag 1470
Db 1783 AATCACCAAAAGAAACAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1842
Qy 1471 tatgaagaagctgtgaagcagatctatgatccaaatcaacaataacatcgactgacc 1530
Db 1843 TATGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1902
Qy 1531 aactcgagaagaatgagaggaagaacggtactctcaacaagtgagagaacatgacacacat 1590
Db 1903 GTCTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1962
Qy 1591 aatgccttgatcctatgagaatctcaagcatatcttgagaacatcgaccacaagcctt 1650
Db 1963 AATTAATTTTATCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2022
Qy 1651 aaglatatgaggaactctctcgcgagacatgltgtgagaagaacataaaglatatc 1710

Db 2023 TCATATCTTGAAGATTATCTTTAAGAAAAAGAAATTTCTGAATAAGATTTTAATCATATAT 2082
Qy 1711 aagaatcctcaatgaatgaatgaagaacgagatcgagacgctgttgagaacatlaagaag 1770
Db 2083 TATACTTTGAATAACTGCGCTCGAAGCTGATTAATAAATAATTAACAGAGAAATTAAGAGT 2142
Qy 1771 gatgaagacagctgtgttgagaagaagatlaacaagaacgaanaataacagatgaagaag 1830
Db 2143 AGTGAACAATAATCTCTGAATAAATAATTTTAAGAGCTTAACACATTCACGAAATG---CT 2199
Qy 1831 atcctgaggtctccgatatgtttaagtlcaagtlcgagaagtgctccatgaagaag 1890
Db 2200 TCTTTAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2259
Qy 1891 atgatgacctcaagaagaccatcatctcttgagaagaacggtgggtttaaacataata 1950
Db 2260 ATGAGAGCTTAAGAAATAATGAATTAATTTAATAAATAATGAACAATTAAGATTAATTT 2319
Qy 1951 catgtgccgaatgatltaagcagagagaataagcaggaaccatacctcatcgtaact 2010
Db 2320 CATGTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2379
Qy 2011 aagaagaatagacaacatgaagtgltcatgcccagaagtcgagagcgtgacacga 2070
Db 2380 AAAAAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2437
Qy 2071 gagaagaagaacatlaaacatgaagagcagatcagatlaactcgagcctccacagaaga 2130
Db 2438 AAGAACAGCTGCTCTTAACAAGTATTAACAACCTTTATGTTGACGAAAGCAACACTG 2497
Qy 2131 gagataaccgagacaggtctaccacaagcccgacaacagcggtctgagcctcgaaagc 2190
Db 2498 AAGAT-GGGGGTCACTCCACACACACATTAATCCCATAGAGAGAAAGAAAGATTAACGAA 2556
Qy 2191 gatagcgtgcaagctcaagcacaagaagcagaagcagacagcctccagtgccagtgccc 2230
Db 2557 GAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2616
Qy 2251 gtccagaaggtlaagcgtcaagtgctcaccacagcgtcctggaataaagaacgag 2310
Db 2617 CCAAAAGAGTAAGATGTTGAAATAATTAATTAAGAACTTAAGAGTAAGAGTAAGAGTAAGAG 2676
Qy 2311 aatgcaagaacatgagacatctgagaagcgtctatgagctcgtgaatlatcctacatc 2370
Db 2677 GCGTTGACAAAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2736
Qy 2371 tgcacaataatlatcctcgtctcctcaagcactatgaacgagaagatcttaacagtac 2430
Db 2737 TGTCTAATAATTAATTTAGTATCAAACTCTGATGAGCAAAATTAATTAATTAATTAATTAAT 2796
Qy 2431 aagataaccaagaagagagagaaagtaaacgtctcctctgtgtatccacgagacgtgttc 2490
Db 2797 AATCTTACTCCAGAAAGAAAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2856
Qy 2491 aatatcagaacaacatcccgatlatgctatgtctgtagtaagcctcaacatctctc 2550
Db 2857 AATATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2916
Qy 2551 tctcaactgtctatgagatatatgagaagagatggtctgcaacgtgtlataaactcaaa 2610
Db 2917 CAACATCTCTTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2976
Qy 2611 gacaacgaagaatlaagaacctctcgaggaagaactaagaaggtctc----- 2657
Db 2977 GAGGAAATCAATCAAAAAATTAATTAAGAGCAAAAAACAAATTAATTAATTAATTAATTAAT 3036
Qy 2658 -----caactctgtaaacatctctctc 2681
Db 3037 ACATCCAGTCTGGAATTAACAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3096
Qy 2682 cagctccatgcaacacactgtctctcaacactcaagaa----- 2720
Db 3097 CAAAACCAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3156

QY 2721 -----gccgaagtgcgcgtacgaacgacctctcactcgcacacttaactca 2775
 Db 3157 GGCCTCGCTGTAGTTGAGAAAGTCATGATCCCTTAACGATATTGCTATTAGTAACGAT 3216
 QY 2776 ctgaacactgttgagacacactcgtctcgcgcgaagaataagaact---ctacccaaga 2832
 Db 3217 TTGAAAGGTATGTGTAGTCTCTTAATCTTGAAATAAAACCTAAATACCTAATTCATTA 3276
 QY 2833 cttaattgacaagaactcgtccgcgaactcttaacgaagaactgaagaacgcgcgaca 2892
 Db 3277 ACCATTTCACACACAGATGGAAAAATTTTATGAGAAATTTTAAAAAATATGATAC 3336
 QY 2893 tctctatacgaagcttcactaactcgtgaaatcaactaagcgcgataatcaactcct 2952
 Db 3337 TATTTTAATGATGATATCAACAACTCGTAAATCTAATTCAAAATTAATACAGCTTG 3396
 QY 2953 aacgatgaatctcaacgtaagaagctgcgaagggagacataactaactgaagaacacg 3012
 Db 3397 ACCGAACACACAAAA--AATGCTTAATATGATGAATTAATAAATTAAGATACCTTTA 3453
 QY 3013 caactgaactcgcacctgtacacaagaactcaaacctgaacctgcgaagactcctgcgaag 3072
 Db 3454 CAGTTATCATTTGATTTATATATATATTAATTAATTAATTAATGATGATTTATTATAG 3513
 QY 3073 aagaagacagctgcgcaagtataagatgcagatcaagaagctgactcgtcgaagcgag 3132
 Db 3514 AAAAAAGAACTTGCCACAGACAAATGCAAAATTAATAAATTAATAAAGAACAA 3573
 QY 3133 ctggaagaagaactcgaactcgaactcgaacatcgaaacgtaactgaagaactcgaag 3192
 Db 3574 TTAGATTCATAATTTGAATTCATTAAATACCCCATTAATGATTAACAAAACCTTTCTGT 3633
 QY 3193 tctctcaacaagaagaagaagccgagatcgccgagacagagaacacactcgcgaagacac 3252
 Db 3634 TTCTTTAACAAAAAAGAGAGCTGAATATGACAAACGTAACATTAAGAAAAACA 3693
 QY 3253 aagatctctctcaaacactaacaagcctgcgtcaaglatataatgycgagctctcct 3312
 Db 3694 AAAATTTTATTTGAACATTAATAAGACCTGTTAAATATATTAATGATGATATCTCCA 3753
 QY 3313 ctggaagaactcctcgcgaagagactcgcgaacgcgagagataactcgaacccgcgaac 3372
 Db 3754 TTAAAACTTTAAAGTAAATCAATTCACAAAGAAATATTAATTCACATTTAGAAAA 3813
 QY 3373 tctcaagctcgtctcgaactcgaagcaagctgaaggaacactcgcgaagaag 3432
 Db 3814 TTTAGAGTATTAAGTAAATAGATGAGAAACCTCAATGATTAATTTAGAGAAAGAA 3873
 QY 3433 aagctgaactcctcctcgaagcgcgactcgaactcgaactcgaactcgaagaatcatt 3492
 Db 3874 AAATTTCTTTCTTATCAAGTGATTAACATTAATTAATTAATTAATAAAGAAATTA 3933
 QY 3493 aagaagaagaactcgaacgcggaagtaagccgaagaagaataatacaagcgtgaatacgca 3552
 Db 3934 AAAAATAAATAATTAATACAGGTAATTTCTCCAGTGAATAATATAGAAAGTTAAGAAAGCT 3993
 QY 3553 ctggaactctcaagaagaactcctcgcgtgaaggaacagatgctgcacactgcygtc--- 3608
 Db 3994 TTAATAATCTTACGAAAAATTTTCTCCAGAAAGAAAGTTACACAGTGTGAACCTCA 4053
 QY 3609 -----tgaactgctccgcgacacactgcgaactgcgaactcgaactcgaactcga 3641
 Db 4054 CAACCATGATTAATTCATCTCCATTAATCTGTAAGGTAAGTGTGATTCAGATCCACA 4113
 QY 3642 acttaagaag-----cctgactatcactcgtgcgagccgaacaaataactaac 3693
 Db 4114 AAGAAGAAACACAAATPACCACTTCAAGCTCTTTTATTAACAGAAATTTACAAAGTAAGTA 4173
 QY 3694 acatcccaagaacgctgcgaatgagctgcgaatgagctcactgctcactcgcgcgag 3753
 Db 4174 CATTTCAAAATTAATGACGAGAAAGATGATTCCTTACTGTGTATTACCATTTTGGAGAA 4233

QY 3754 aagcgaaggaactaagaatgaactcgcgcaggtggtltaacccgtgagcgtctcactcc 3813
 Db 4234 TCCGAGATTAATGACGAATAATTTAGATCAAGTATCAAGTATCGAGAAC---AATATCTGTC 4290
 QY 3814 gtagatgaataactctgctccaaatactcgaagaacgaatacgaagctgctctatcgaacact 3873
 Db 4291 ACAATGATTAATTCCTCTCAGGATTTGAATAATGAATTAATGATTAATTAATAACCT 4350
 QY 3874 ctgcgcagcgtctatagctctcgaagaacagctgcgagaataaactgactgaactcaat 3933
 Db 4351 TTAGCTGAGTATATGAGAGCTTAATAAATAAATAATGAATAAATAATTTTACATTAAT 4410
 QY 3934 gtcaacgtgaaggaactctcgaacgcgcgtcttaataagaaggaataatcgaagaacgc 3993
 Db 4411 TTAATAATTAACGATATCTTAATTAATTCAGCTCTTAATAAAGAAATTAATTTAGATGA 4470
 QY 3994 ttagaagagcactgtactccctataagaacactgaactcctcactgaactgctgaagac 4053
 Db 4471 TTAGAACTGATTTTAATGCAATTTTAACATATATCTCAAAATGATATCATTTATGAAGAT 4530
 QY 4054 ccaataaagctcctcaataaagaagaaggaataaattctcgtctgaactaactatc 4113
 Db 4531 TCATTTAAATTAATTAATGATTCAGAACAAAAACACACTTTTAAAAAGTTACAAATATATA 4590
 QY 4114 aaggaactcactgcacacgcgatatcaattcgcgtcaatgagtgcgtggtattcaagaatc 4173
 Db 4591 AAGAATTCAGTAGAAATGATATTAATTAATTTGACAGAGAAAGTAAATTAATTAATGAAG 4650
 QY 4174 ctgagcgaataatacaagctcgtacactcgtactcattaaagaatatacaagaataag--- 4230
 Db 4651 GTTTAGCGAAATATTAAGATGATTTAGATTCATTTAAAAAGTTTCAAGAAGAAAG 4710
 QY 4231 -----caa 4233
 Db 4711 GAGAAATTCATCATCACACACACACACACCTCCCTCACAGCAAAAAACAGACGAACAA 4770
 QY 4234 ggcggaatgaaataatctgcctcctcgtgaataatcgaacactgcgaagaagc 4293
 Db 4771 AAGCAAGAAAGTAAAGTTCCTTCATTTTAAACAAATTTGAGACCTTAATTAATTAATCTTA 4830
 QY 4294 aagcgaataatcgaactcctgcgaactcctgcgaagcgaagctcctcaactaactaag 4353
 Db 4831 GTTAATTAATTTAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4890
 QY 4354 gagaagagcaatgctgaagcttaataatcaagagcgtgaactcctcaaaactcaagaac 4413
 Db 4891 GAAAAAGATGAAACACATGTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4950
 QY 4414 aagctgcgacttcaagaataaatacaattcgtcgaatgcaagctcgtctacgat 4473
 Db 4951 AAATAGATCTTTTAAACCAATTAACGACTTCGAAGCATTAATTAATAATTAATTAATTAAT 5010
 QY 4474 tataacacacacactcctcgaacgaactcgtctcaacgcaagctgctgcgaacactc 4533
 Db 5011 GATACGAAAAAATATGCTTGCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5067
 QY 4534 gcccaaacagctgcgaactcgtcgaacgcggaacactcgaaggaactgctgaacactcc 4593
 Db 5068 CTAATATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5127
 QY 4594 cagacacaaatgcgtgaagaagaagctcccccgaataagagcgtgttgcgaactcgcgac 4653
 Db 5128 CAACACCAATGCGTAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5187
 QY 4654 gacgcgaagagctgcgaactcctcgaactcaacaaagaagaagataagctgcgtgag 4713
 Db 5188 GAAGAAGAAAGTAAATGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5247
 QY 4714 aaccacaacactcgtcaatgaaacaatgycggtgtgtaacgcgataatgcaatgcaatgca 4773
 Db 5248 AATCCAAATCCCTACTTGTAAAGAAATTAATGATGATGATGATGATGATGATGATGATGATG 5307
 QY 4774 gaggaagaacgcgctcactaaggaagaataactcaactgcgaatgctgaagccgactcc 4833

Accession	Sequence	Position
Dd	GAAGAAGATTGACGTACGACCGAAGAAAGAAATGCAATGTACTTAACCTGATTTT	5367
Oy	4834 tatccactcttcagacgggattttctgctccagctctaattccctgggacatccctctctg	4893
Dd	5368 TATCCACATTTTGGATGGATTTTCTGACAGTTCCCTTAACCTCTTAGAATATATCTTCTA	5427
Oy	4894 ctgataccctcaatgctatccctgaacgcttcaatcctaatatgatcgatg	4939
Dd	5428 TTAAATCTCATGTATAATTATACAGTTCTTAAAAATCTTAGG	5473

RESULT	17
PfAMSP806	
LOCUS	PfAMSP806
DEFINITION	1636 bp DNA linear INV 03-FEB-1999
ACCESSION	PfAMSP806
VERSION	P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.
KEYWORDS	D13357
SOURCE	D13357.1 GI:391784
REMARKS	EgF-like domains; major merozoite surface protein precursor. Plasmodium falciparum (isolate 806) merozoite, DNA.

BASE COUNT	696 a	199 c	239 g	502 t
ORIGIN	Chromosome 9.			

Query Match	18.4%;	Score 910.4;	DB 3;	Length 1636;
Best Local Similarity	72.4%;	Pred. No. 5.7e-190;		
Matches 1181; Conservative	0;	Mismatches 451;	Indels 0;	Gaps 0;

OY	3308	ctctctcgaagacactctccgagagagatccacgacgagatatactacgacgcgcg	3367
Db	1	ctcccttataaactcttttaagtaagaaatcaattttataacacaaatattttatgccactttag	60
OY	3368	agaacttcaagctcgtctcctaagctcgaaagcgaaagcttgaaagacacaacttgaaactcgagaa	3427
Db	61	aaaactttaaagctatttaagtaaaatttagaagaaaatttaaaagctattatttttaatttagaaa	120
OY	3428	agaagaagcttaagctactctctcagcgagactgatactcgtgcgcgagctcgaagag	3487
Db	121	agaaaaaattttctactttttacaaagagattatcatatttttaatttcctgatttaaaagag	180
OY	3488	lcaatlaagaacaagaactataccggcaatlaagcccaagcagagataatacagaactgaa	3547
Db	181	taattaaaaaattaaaaattttatcacagatattctccaagtaaaaattatagcgatttaaca	240
OY	3548	acgcacatggaatcttcaagaagttcctgcgtcgaagaagcaatgtgcacactggtggt	3607
Db	241	atgcatttttaattttttttacaaaaaattttctcccaacaaagaaacaatgcttgcaacagtgtgaa	300
OY	3608	ctgaatctcgctccgacacacacttgagcagttcctaaccctaagaagctgcattactacg	3667
Db	301	gtmaaaagtgatccgacacattatgaaacaaagtcacaaacaaagaaacacagatcaactcatg	360
OY	3668	tcgagcgcgagttccaatacaatataccaacatctcaagaagctgaagatgagttcgatgcg	3727
Db	361	tagagacgagattcttaacacataatcaacacatcacaaaaatttcgatgatgaagtagatgcg	420
OY	3728	lcaatcattgtgcctactcttcggcgagagcgaagagagactacagactcgacccgcgactg	3787
Db	421	taattcatatgatactatttttggaagaaatccgaacaaagatttatgatgttttagacaaatag	480
OY	3788	tcaccgatgaggtcgtcactcctctcgttgattgataacatctgtccaaaatcgagaagc	3847
Db	481	ttaacagcgagaaacggtatgcctttccgtaatttgatatacctttcttaaaatttgaaaatg	540
OY	3848	aatacgagtgctctatacttcaaaacctctggcagcgctcatagttctctcaagaacagc	3907
Db	541	aattatgaggttttatatttttaaaacgttttagacagcttttatagaagtttttaaaaaacat	600
OY	3908	ttgsgaataacgtlgaagacttcaatgtlcaagcgtgaagacatcttgaacagcgctta	3967
Db	601	tagaaaaattaacgtttacacttttaaaattgtaattgttagcatttttaaaattccacatttta	660
OY	3968	ataagagaaaatcttcaagaacgctcttggagcgagacttgctccctataagaagcctga	4027
Db	661	attaaacgtgaaaaattttcaaaaaattgttttagaaatcacagatttaatttcctatrtaaagtttta	720
OY	4028	ccctcctaactacgtgtlgaaggaccacatacaagttcctcctaataagaagaagagata	4087
Db	721	catcacagtaattatgtgttcmaaaagatccattatataattttttatataagaaaaaacagtta	780
OY	4088	aattcgtctcagttacaactatatacaagagcccatcgacaccgatatacattctcgta	4147
Db	781	aattctttagcaggttatatttatatttagcatttcattatgcgatatttaatttttgcaa	840
OY	4148	atgattgtcgtgggtlatacagaatccctgaagcgaanaatacaagctcgaacttgactta	4207
Db	841	atgatgtcttcttgcatattttataaaaaattttatccgaaaaaaatttaaacagtttttagattcaa	900
OY	4208	ttaaaaagttatacaagataaagcgaagcggagatgaiaaanaatatctgccttcctcgta	4267
Db	901	tttaaaaaattatattacacgaacaaacaaagctgaanaatgacaaatccttcccttttttaaca	960
OY	4268	acatcgaaacctgtataagacagltgaagcacaanaatcgacctcttcglaatcaacctg	4327
Db	961	atatttgagaccttatataaaacagtttaagcattataaatttatatttttttaatttcattttag	1020

OY	4328	aggccgaagcttcctaacttaacttaagaagaagcagatgtagaaatttaaatcaaacgc	4387
Db	1021	AAGCAAAAGTTCTTAATATTATACATATGAGAATCAAACTGAGAGTTAATAATTAAGAAC	1080
OY	4388	tgaactacctcaaaaacataccaagaacgtggcgagatlccaagaataatacaatttcg	4447
Db	1081	TTAATTAATTAATAACAATTCACAGCAAATGGCAGATTTTAAAAAAAATAACATTTTG	1140
OY	4448	tcggaattcgaccctgtctacgcgattataacaacaacatctcccgaccaagtctgt	4507
Db	1141	TTGGAAATTCGATTTATTCACACAGATTATACCCTAATAATTAATTCAGCAAAAGTTCCCTA	1200
OY	4508	ccactggaatggigtgtcgaanaactgccaaaaacagagcgtgacgatctgctgacgga	4567
Db	1201	GTAACGATGATGTTTTGAANAATCTGTCTAATAACCGTTTATCTAATTTACTTGATGAA	1260
OY	4568	accgtgcaagtgatctgcaacatctcccagaacacatgctggaagaacagtgccccaga	4627
Db	1261	ACTTGCAAGTATATGTTAAACATTTCACACACCAATCGTAAAAACAATGTCACAAA	1320
OY	4628	aatgacgctgttccaagcatctgacgacgacgacgacgtgacgtgctctcgaactaca	4687
Db	1321	ATTCGATGATGTTTGCAGCATTTGATGAAGAGACAGATGAATGATTTATTAATTTACA	1380
OY	4688	aacaagaagsgataaagtgtggtgagagaccacaacccctactcgcaatgaaacaatggcg	4747
Db	1381	AACAGAGGATGATTAATATGTTGTAATAATCCAAATCTCTACTTTGACGAAAAATAGTG	1440
OY	4748	ggtgtgacgcgcatctcaaatgacacgagagaagacgacgctctcaacggaataaataca	4807
Db	1441	GATGATGATGATGCCAAATGTAACGAGAGAGTTGAGTTAGTACCAACGAAAAAATATCA	1500
OY	4808	cattgagtgtaactaaagccgactccctacatccactctcgacggagatttttgtccagct	4867
Db	1501	CATGATGATGATTAACAACTGATTTATTCACACTTTTCGATGATTTTTCGAGTTTCT	1560
OY	4868	ctaatttcctggatctccctctctctgctgtagtcccatcgatccgtaagatctatct	4927
Db	1561	CTAATCTTTTAGAATATCATCTTTATTAATCACTGATTTATTAATTAACAGTTTCACTT	1620
OY	4928	aatatgcatg 4939	
Db	1621	AAAAATGTAG 1632	
RESULT 18			
PFA MSP808			
LOCUS	PFA MSP808	1636 bp DNA linear INV 03-FEB-1999	
DEFINITION	P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.		
ACCESSION	DJ3358		
VERSION	DJ3358.1	GI:391788	
KEYWORDS	EGF-like domains; major merozoite surface protein precursor.		
SOURCE	Plasmodium falciparum (isolate 808) merozoite, DNA.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 1636)		
TITLE	Jongwulwies, S.		
JOURNAL	Submitted (05-OCT-1992) Somchai Jongwulwies, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology, 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)		
REFERENCE	2 (bases 1 to 1636)		
AUTHORS	Jongwulwies, S.; Tanabe, K. and Kanbara, H.		
TITLE	Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP) of Plasmodium falciparum from field isolates		
JOURNAL	Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)		
MEDLINE	93295445		
COMMENT	Submitted (05-OCT-1992) to DDBJ by: Somchai Jongwulwies Department of Protozoology		

D	b	1	CTCCATTAAAAACTTTAAGTGAAGAAATCAATTCGAAACAGATATATATGCAAGTTAG	60
Q	y	3368	agaacttaagtcctcgtcttaagctcgaaagcgaaacttgaagacaaccgaaaccggaga	3427
D	b	61	AAAATTTTAAAGTATTAAGTAAATTAACAAAGAAAATTAAGATTAATTTAAATTTTGA	120
Q	y	3428	agaagaagctagaactactcctctagtcgagactgcatcacttgcgcgagctcaagaga	3487
D	b	121	AGAAAAATTTATACACTTATCAAGTGGATTTACATCTTAATTTGGTGAATTTAAAGAAG	180
Q	y	3488	tcaataagaacaagaactacaacgcggaatagcccaagcgagaataatacagaagtgaa	3547
D	b	181	TAAATAAAAAATTAATTAACAGATTAATCTCAAGTAAAAATATATAGATGTTAAACA	240
Q	y	3548	acgacgtgaaacttacaagaagcttcgtgcgaagaagacagatgctgcacttgygct	3607
D	b	241	ATGCTTATGAATCTTTACAAAAAATTTCTCCAGACAGAACACAGTGTTCACACGTTGTAA	300
Q	y	3608	ctgaaacttgcgtccgacacacatgtagcgagcttcaaccttaagaagccctgactactacg	3667
D	b	301	GTGAAGAGTGATTCGACACATTAAGAACAAAGTCAACCAAGAACACAGACTCAACTCATG	360
Q	y	3668	tggagccgagatccaataaataccaactctagaagctgaagatgagtgcgatagc	3727
D	b	361	TAGGAGCAGAGCTTACACATTAACAATCAACATCAACAAATGTCAGTGTGAAGTAGATGAGC	420
Q	y	3728	tcaatattgtgctcacttctcggcgagagcgaaagagactacgatgacctgcgcagctg	3787
D	b	421	TAAATCATGTACCTATATTTTGGAAATCCGAAGACATTAATGATGTTTAGACAAATAG	480
Q	y	3788	tcaacgggagagcgctgaacctccctcgtgagtataaacttctgtccaaaaatcsgaagc	3847
D	b	481	TAAACGGGAAAGCGTAACTCTTCGTAATTTATATACACTCTTCTTAAATTTGAAATG	540
Q	y	3848	aatacgaagtgctatctatctgaacctctggcagcgctcatalagtgctctcaagaacagc	3907
D	b	541	AATATGAGCTTTTATATTTAAACCTTTACAGGTGTTATGAAGTTAAAAAACAT	600
Q	y	3908	tggagaataacggtgataccttcaagtgcaagtgaaagacatcttgaacagcgcttca	3967
D	b	601	TAGAAAATTAACGTTATATGACATTTAATGTTAATTTAAGATATTTTAAATTTACCATTTA	660
Q	y	3968	ataagaaggaataattccaagaagctcttggagcgagacttgatcccttaagaacctga	4027
D	b	661	ATTAACCGTAATTTTCAAAAATGTTTGAATTCACATTTAATTCATATTAAGATTTTAA	720
Q	y	4028	cctcctctaactacgtgtgcaaggaaccaatacaagttcctcaataaagaagaaggagta	4087
D	b	721	CATCAAGTAATTAATGTTGTCAAGAGATCCATATTAATTTCTTATTAAGCAAAAACAGATA	780
Q	y	4088	aatttcgtctaagttacaactataccaagaagcccgatgaagccgatatcaatttcgta	4147
D	b	781	AATTTTAAAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTGA	840
Q	y	4148	atgatatgcttggggtattacaagaactccttggcgaaataatacaagttgacttgcactta	4207
D	b	841	ATGATGTTCTTGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	900
Q	y	4208	ttaaaagtataatacaagaataagcaagcgagagatagaataatcttgccttccgtaata	4267
D	b	901	TTTAAAAATATATTAACGACAAACAAAGCTTAATTAAGAAATTAACCTCTTTTAAACA	960
Q	y	4268	acatcgaagaccctgtacaagaagtgaaagcaaaaatgcacccttcgtaattcaaccg	4327
D	b	961	ATATTGAGACCTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1020
Q	y	4328	agggccaaggtctcaactatacttcagcaagaagcaatgttggaaagttaaatcaagaagc	4387
D	b	1021	AAGCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1080
Q	y	4388	tgaactactcttaaaaacaataccaagcaagcagctgtcgatcttaagaanaatacaattcg	4447
D	b	1081	TTTAATTTACTTTAAAAACAATTCACACAAATTTGGCAATTTTAAAAAAAATTAACAATTTTG	1140

Oy	4448	toggatagtcaggactgctcaccgatttataaccaaaaacatctccgcgccaaagtctgt	4507
Db	1141	TTGGATTGCTGCAATTAACAACAGATTATAACCATTAATAACTTATTGAACAAAGTTCCTTA	1200
Oy	4508	coactcgcatgtgtctcgaaaaaacctccgcaaaacagtgctggaacaatcgtcgcagggca	4567
Db	1201	GTAACAGGTATGCGTTTTTGAAAAATCTTGCTAAAACCCTTTATCTAATTACTGTGATGGA	1260
Oy	4568	acctgcagggcactgctggaacatctcccagaaccaatgctgtaagaagaacagtgccccaga	4627
Db	1261	ACTTGCAAGGATATGTTAAATCAATTTACACAACCAATGCGTAAAAAAACAATGTCAGAAA	1320
Oy	4628	atagcgctgttttcagsgaatcgcagcagcgcggaagatgcaagtgctccctgcgaactaca	4687
Db	1321	ATTCGCGATGTTTCCAGACATTTTAGTATGAAGAAGAAAGAAATGTAATGTTTATTAAATTTACA	1380
Oy	4688	aacaaagaagsgataagtcgcltggaaaccccaaacctcactgcaycaatgaaacaaatggcg	4747
Db	1381	AACAAGAAGGTGATTAATGCTGTGAATAATCCAAATCCTATGTTAACGAAATAATATGNGH	1440
Oy	4748	ggltgtgaagccgatgtctaattgacccggggaagaagaagcgctctaaggaagaanaatcca	4807
Db	1441	GATGGAAGCGAGATGCCAAATGTPACCGAAGAAGATTCAAGTAGCAACGAAAGAAATCA	1500
Oy	4808	cattgagatgtactaagaaccgcagctcctaaccacttcgcagcggatllttgctccagct	4867
Db	1501	CATGGAATGTACTAATACCTGATTTCTTATCCACTTTTGATGATGTAATTTTCTGCATTTGCT	1560
Oy	4868	ctaattccctgggaatcctcctcctgctgatacctcctgactgatacctgatacagctcaatc	4927
Db	1561	CTAATCTTTTGAAGATATCATCTTTATTATATCTATATATTAATATATACAGTTTCATTT	1620
Oy	4928	aatatagtcgatc 4939	
Db	1621	AAAAATGTAGS 1632	
RESULT 20			
PfPAMSP837			
LOCUS		1636 bp DNA linear INV 03-FEB-1999	
DEFINITION		P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.	
ACCESSION		D13362	
VERSION		D13362.1 GI:391806	
KEYWORDS		EGR-like domains; major merozoite surface protein precursor.	
SOURCE		Plasmodium falciparum (isolate 837) merozoite, DNA.	
ORGANISM		Plasmodium falciparum	
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
TITLE		1 (bases 1 to 1636)	
JOURNAL		Jongwuiwies, S.	
REFERENCE		Direct Submission	
AUTHORS		Submitted (05-OCT-1992) Somchai Jongwuiwies, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel.:0958-47-2111(ex.3747), Fax:0958-47-6607)	
TITLE		2 (bases 1 to 1636)	
JOURNAL		Jongwuiwies, S., Tanabe, K. and Kanbara, H.	
MEDLINE		Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates	
COMMENT		Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)	
		93295445	
		Submitted (05-OCT-1992) to DBJ by:	
		Somchai Jongwuiwies	
		Department of Protozoology	
		Institute of Tropical Medicine	
		Nagasaki University	
		12-4 Sakamoto-machi	
		Nagasaki 852	
		Japan	
		Phone: 0958-47-2111x3747	
		Fax: 0958-47-6607.	

FEATURES		Location/Qualifiers	
source		1..1636	
		/organism="Plasmodium falciparum"	
		/db_xref="taxon:5833"	
		/dev_stage="merozoite"	
gene		1..1622	
		/gene="MSP1"	
		<1..1622	
CDS		/gene="MSP1"	
		/codon_start=3	
		/product="major merozoite surface protein precursor"	
		/protein_id="BA02623.1"	
		/db_xref="GI:391807"	
		/translation="PLKTLSEESIQEDNYASLENEFKVLSKLGKLNLEKKKLS YLSGJHHLIAELKEVTKNKNYGNPSVNTDYNNALESYKKFLPEGTAVATVSES GSDTLEOSOPKRKPSATHVCAESMTITTSQVNDVDEVDITVPIFGESSEEDYDLGOVY TGEAVTPSYIDNLISKIENEYEVLYTKPLGAVRSKJQJLENNMFTFNVAKDILNSR FNKRENFKNVLESDDLIPYKDLTSSNVVKKPKFLEKREKDFLSSVNYIKDSIDDI NFANDYLGYYKIKLEKKSDLSIKKYNKOENKEYLPLNNIETLYTVDKIDL FVJHLEAKVLYNTYERKSNEVKIKELNYLTIDKLADFKKNNFVCIADSTIDYNN NLTKFLSTGMFENLAKTVLSNLDGNLGMIDNISQHOCKVKOCPONSGCFPHLDER EERCKLNTYKQBGDKCVENPNCNENNGSCDADACTEEDSGSKKIKTECTKPS YPLFDGIFCSSNPLGISFLILMLILYSFI"	
BASE COUNT		694 a 199 c 239 g 504 t	
ORIGIN		Chromosome 9.	
Query Match		18.4% Score 908.8; DB 3; Length 1636;	
Best Local Similarity		72.3% Pred. No. 1.3e-189;	
Matches 1180; Conservative		0; Mismatches 452; Indels 0; Gaps 0;	
Qy	3308	ctctctgaagactctctccgagagagagcaccgagcgaagataactagccagccctcg	3367
Db	1	CTCCATTAAAACTTTAAGTCAGAACATCAATTCACAGAGATTAATTCACAGTTTAA	60
Qy	3368	agaactcaagtcctcgtcctaagctcgaagcgaagctgaagacaactgaagctggaga	3427
Db	61	AAAACCTTTAAAGTATTAAGTAAATTTAGAGAGAAATTAAGCATTAATTTAATTAAGAA	120
Qy	3428	agaagaagcctcagctacccctctcagagagctgcataccctgcagcgcagctgaaga	3487
Db	121	AGAAAAATTAATTCATTTATCAAGTATCAATTCATTTAATTCATTTAATTAAGAAC	180
Qy	3488	tcattagaacaagaactacacacgcggaatagcccaagcgaagataatacagaagctgata	3547
Db	181	TAAATAAAAATTAATAATTAATTCAGGTAAATTCCTCAAGTGAATTAATTCAGATTTAACA	240
Qy	3548	acgcactgaagactcttacaagaagctcctgcctgaagacaagatgctgcacatgtggt	3607
Db	241	ATGCATTAGAAATCTTACAAAAAATTTCTCCAGAAAGACGATGTTGCACAGTTGTA	300
Qy	3608	ctgaatctgctccgaacacactgagacagctcacaactaagaagcctgcatactcatg	3667
Db	301	GTCGAAATGGATCCGACATTTAGAACAACTCAACCAAAACCAAGCAATCAATCATG	360
Qy	3668	tcgagacgcagctccaatcacatctcagaacgctgcagcagctgaagctgagctgagc	3727
Db	361	TAGAGACAGAGCTCTAACACATTAACAACAAATGTCAGATGATGAATGAATGACG	420
Qy	3728	tcacatctgctcctcctcgc	3787
Db	421	TAAATCATAGTACCTAATTTGGAGCAATCCGAAAGAAATTAATGATGATTTAGCAAGAG	480
Qy	3788	tcacgcgtgagcgtgctcaactccctcgcgtgattgataacatctctccaaaatcgaga	3847
Db	481	TAAACGAG	540
Qy	3848	aatacgaagtgctctacatcgaacactcggaagcgcctataggtctcctcaagaagaagc	3907
Db	541	AATATGAGGTTTATATTTAAACCTTTAGCAGGTGTTTATAGAGTTTAAAAAACAAT	600
Qy	3908	tgagagaataacgltgacactcgaatgcaacgtgaagaacatctgaacagccgcttca	3967
Db			
Db	601	TAGAAAAATACGTTATGACATTTAATGTTAATGTTAAGATATTTTAAATTCACGATTTA	660
Qy	3968	ataagagagaataatccaagaagctcttgagagcagcttgatccctcctataaagactga	4027
Db	661	ATTAACGAGGAAATTTCAAAAATGTTTATGATCAGATTAATTCATATTAACATTTAA	720
Qy	4028	ccctcctcaactagctgtcgaagagccatcaagctccctcctcaataaagaagaagagata	4087
Db	721	CATCAGTAATTAATGTTCTGTCACAGATTCATTAATTTCTTAATTAAGAAAAAGAGATA	780
Qy	4088	aattcctcgtcagttacaacatataccaagaagctccatcgacacagcatatacattcgcta	4147
Db	781	AATCTTAGACGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	840
Qy	4148	atgctgctgctgaggtatatacagaagctcccgagcgaagaaataaagctgacttgactcta	4207
Db	841	ATGATGCTTCTGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	900
Qy	4208	ttaaaagatatacaacgataaagaagcgaagctgaagaaatatactgcctcctgata	4267
Db	901	TTAAAAATATATATACAGACACACAGAGTGAATTAATTAATTAATTAATTAATTAATTA	960
Qy	4268	acatcgaaacccctgtacaagaagctgaagcgaagaaatcgaccctcctgtaactcgtg	4327
Db	961	ATATTGAGACCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1020
Qy	4328	aggccaagctcctcaactatctcagagaagaagcaatgtggaagttaaatccaagagc	4387
Db	1021	AAGCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1080
Qy	4388	tgaactccctcaaaacaatcccaagaagcgtgcagatttcaagaagaataatacaatttcg	4447
Db	1081	TTAATTTACTTAATAAACATTTACAGCAATTTGCGAGATTTTAAAAAATTAACATTTTCG	1140
Qy	4448	tcggaattgcagaccctgtcctacgaattataccaacaacatccctcgaaccaagttcgt	4507
Db	1141	TTGGAATTTGCTGATTTATCAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTA	1200
Qy	4508	ccactgcatgctgtcttgaaagaactcgcgaagaagaagctggaagatctgctgaagcga	4567
Db	1201	GTCACAGTATGTTTGGAAAAATCTTCTTAACACCGTTTATTAATTTAATTTATGATGAA	1260
Qy	4568	accctgcagctcgtgaacatctcccgacacagctgcgtgaagaagaagctcccgaga	4627
Db	1261	ACTTGCAAGGATTTTAACATTTTACACACACCAATGCGTAAATAAATAATGTCACAAA	1320
Qy	4628	atagcgctgtctcagcatctggaagcgcgcgaagaagctgcaagctgctcctcgaactaca	4687
Db	1321	ATTCCTGATGTTTACACATTTAGATGAAGAAGAAATGTAATGTTTATTAATTAATTAACA	1380
Qy	4688	aacaagaagaagataagctgctggaagaaccgaacccctcactcgaatggaagaagaagc	4747
Db	1381	AACAAGAGAGGATTAATTTGTTGAAAAATCCAAATCCCTACTTGAACGAAATTAATGCTG	1440
Qy	4748	ggtgtgacgcgcgaatgctcaaatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	4807
Db	1441	GATGTGATGCGAGATGCCAAATGTCAGAGAGAGATTAACAGTACGACAGGAAAGAAATCA	1500
Qy	4808	catgcagctgactaagcccgacactcctacactccttcgacgaggaattttgtcctcagct	4867
Db	1501	CATGTGATGATTAACCAACCTGATTTATCCACTTTTGCATGATGATTTTTCGAGTTCT	1560
Qy	4868	ctaatttcctggagcatctcctcctcgtgatacccaatgctgcatctgtataagcttaact	4927
Db	1561	CTAATCTTCTTAAGAAATTCATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA	1620
Qy	4928	aatacagcatg 4939	
Db	1621	AAAAAATGTAG 1632	

RESULT 21
PFAMSPn9

LOCUS PFMASPP9 1636 bp DNA linear INV 03-FEB-1999

DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

ACCESSION D13356

VERSION D13356.1 GI:391824

KEYWORDS EGF-like domains; major merozoite surface protein precursor.

SOURCE Plasmodium falciparum (isolate T9/94) merozoite, DNA.

ORGANISM Plasmodium falciparum

REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

AUTHORS 1 (bases 1 to 1636)

TITLE Jongwutives,S.

REFERENCE Direct Submission

AUTHORS Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology, 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)

TITLE 2 (bases 1 to 1636)

AUTHORS Jongwutives,S., Tanabe,K. and Kanbara,H.

REFERENCE Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates

JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)

MEDLINE 93295445

COMMENT Submitted (05-OCT-1992) to DBJ by:

Somchai Jongwutives

Department of Protozoology

Institute of Tropical Medicine

Nagasaki University

12-4 Sakamoto-machi

Nagasaki 852

Japan

Phone: 0958-47-2111x3747

Fax: 0958-47-6607.

FEATURES

source

location/Qualifiers

1..1636

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/dev_stage="merozoite"

1..1622

/gene="MSP1"

<1..1622

/gene="MSP1"

/codon_start=3

/product="major merozoite surface protein precursor"

/protein_id="BAA02617.1"

/db_xref="GI:391825"

/translation="PLKTLSEESIQTEDNVASLENFVLSKLEGLKDNLNEKKKLS YLSGHLHLIAELKEVINKKNVGTGSPVNMVDVNNALSEYKFLPEGTDAVAYSEES GSPVLEPOSPKRPASTHYGASRNITTSQNDDEVDVITVPIGSESEDYDDICQVY TGEAVTPSPIDNILISKINETEVEVLYLKLAEVYSLKQLENNWTFVNVKDLINS FNKRNFENFVLESDLIPIKDLTSSNVVVDYKFLNKRKRDKFLSSVYIKDSIDTID MFADVILGVYKILSEKYSKSDLSIKKYINDROGENEKLPFLNNIETLYKTVNKK IDIFVILAEKVLVNTYEKSNVEVKIKELNLTLOKDLAEFLKNNNVGFIADISTDYNNH NLTKFLSTGWFLENLAATKATVLSLIDGLMGLNLSQHCYKCKPOKSGCFRHLDEFE EECGLILNKRQEGDKCVENPNTCEENNNGCDADAKCTIEBDSGSKGKTIETCEKTPDSE YLPFDGICFSSSENLGISFLILMLILTSFT"

BASE COUNT 695 a 199 c 239 g 503 t

ORIGIN Chromosome 9.

Query Match 18.4%; Score 908.8; DB 3; Length 1636;

Best Local Similarity 72.3%; Pred. No. 1.3e-189;

Matches 1180; Conservative 0; Mismatches 452; Indels 0; Gaps 0.

3308 ctcctctgaagctctcttcgcgagagagatccagaccgagataactcgcgaactcg 3367

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1 CTCCACTTAAAAAAGCTTTAACTGAGAGATCAATTCACAAACAGATATATTATGCGACGTTAG 60

3368 agaactcagaagcttcgctcgaagctcgaagcgaagctgaaggaacaacctggaag 3427

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 61 AAAACTTTAAAGATATTAACTTAATTAAGAGGAAATTAAGATATATTAAATTTTGA 120

3428 agaaagaagctcgaactcctctacgcgaactcgaactcgaactcgcgcgaagc 3487

Db	121	AGAAAAATTATCACTACTTATCAAGAGATTACATCATTTAATTCGTGATTTAAAGAAG	180
Qy	3488	tcattagaacaagaactacacgcgcaatagccacgcagataataacacagctgnaa	3547
Db	181	TAATAAAAAAATAAATAATTAACAGGTAAATCTCCAGTGTAATATACGATGTTAAACA	240
Qy	3548	acgcacgcgnaatcttaagaagttcctgcgtgaagaacagatgctgcacatgtyggt	3607
Db	241	ATGATATTAGATCTTTACAAAAAATTTCTCCCAAGAACGACAGTGTGTGAACAGTTGTAA	300
Qy	3608	ctgaatctgcgtccgcacacactggagagctcaactaagaagcctgcatctacatg	3667
Db	301	GTGAAAGTGGATCCGACACATTAGAACAAAGTCAACCAAGAAACACAGATCACTCAG	360
Qy	3668	tcggagccgagtcaccaataactaccacatctcaagacgtcgacgtgaagtcgatlga	3727
Db	361	TAGAGACGAGATCTTAACAAATTAACAACATCAAAATGTCGATGTAAGATAGATGACG	420
Qy	3728	tcacatctgtgctactatcttcggcgagagcgagaagacacagatgaccccgccagtyg	3787
Db	421	TAACTCATAGTACCTTATVTTTGGAGAAATCCCAAAACAAATATATGATGTATGACAAAGT	480
Qy	3788	tcacggctggagcgtcactacccctccgtgtatataacatctgctccaaatccgaagc	3847
Db	481	TAAACAGGAGAGCAGTACTCTTCCTGATTTATATACATCTTTTAAATTTGAAAG	540
Qy	3848	aatacgaatgctactatcgtaaacctctggcaagcgtctaaagttcctcaagaacagc	3907
Db	541	AATATGAGGTTTATATTTAAACCTTTACAGAGTGTATATGAAAGTTTAAAAAACAAT	600
Qy	3908	tgaggaataacgfbatcacttcaatgtaacgcgtgaagacatctcgaacagccgttta	3967
Db	601	TAGAAAAATACGTTATACACTTTTAATGTTAATTTAAGATATTTTAAATTTACAGATTTA	660
Qy	3968	ataagaagaaatcttaagaagcgtctggagaagcgacttgatccctataagaacctga	4027
Db	661	ATTAACGTAAGAAATTTCAAAAAATGTTTAAATCAGATTTAATCCATTTAAAGATTTAA	720
Qy	4028	ccctccctaactacgttgttcaaggaaccatacagaatcctccaataagaagaagggata	4087
Db	721	CATCAAGTAATTAATGTTGTCGCAAAAGTCCATATTAATTTCTTAATAAGAAAAAGATA	780
Qy	4088	aatttcgtctagtatacactatacagaagccatccatcgacacagcatcaatcttcgcta	4147
Db	781	AATCTTTAAGCAGTTAAATAATTAATTAAGGATTCATATAGTGATTAATAATTTTGC	840
Qy	4148	atgaatgctgagggtatataagaatcctcgagcgaataatacagaatctgcactgctcta	4207
Db	841	ATGATGTTCTTTGGATATTATTAATAATATTATCCCAAAATATTAATCAGTTTAGATTCA	900
Qy	4208	ctlaaaagtalatacaagcgaataagcgaagcgaagataaatactctgcctctccygaata	4267
Db	901	TTAAAAATATATACAGACAAACAAAGTAAATGAGAAATACCTTCCCTTTTAAACA	960
Qy	4268	acatcgaaacccgttacaagacagtggaacgacaanaatcgacctctcgtlaatcacctg	4327
Db	961	ATATTGACCTTATATATAAAACAGTTAATGATTAATTAATTTGTAATTTCAATTTAG	1020
Qy	4328	agggccaaggtccttaactatacttaogagaagcgcaatgtggaagttaaatccaagggc	4387
Db	1021	AAGCAAAAGTTTAAATTTATATACATATTCAGAAATCAAAACGTAGCAAGTTAAATPAAAG	1080
Qy	4388	tgaatcaccctcaaaaacaatccaagaacgctggcagatctcaagaaaaatacaattccg	4447
Db	1081	TTAATTTCTTAAANACATTTCAAGACAAATTTGGCAATTTTAAAAAAAATAACAATTTTC	1147
Qy	4448	tcgaatctgcagacctcttaccgatatalaaccacaacaatctccgaccagaatttcgt	4507
Db	1141	TTTGAATTTTCGATTTATTCACAGAGATTATATACATATATACATTATTGCAAAAGTTCTTA	1200
Qy	4508	ccactggtcatggtgttcgaaaaacctcgccaaaacagtgctgagcaatctgctcgacgga	4567

Db 1201 GTACAGATGTTTGGAAAATCTTGCTAAACCCTTTATCTAATTACTGATGGA 1260

Qy 4686 accgcagggatctgaacatctcccaagccaatggtgaagaacagtgcccccaga 4627

Db 1261 ACTGCAAGGATGTTAAACATTTTCACACACCAATGCGTAAAAAACAATGTCACAAA 1320

Qy 4628 ataagcgctgttcaagcaltcgcagcagcgagaaatgltcctcgaactaca 4687

Db 1321 ATTCTGATGTTTCAGACATTATAGATGAAGAAGAAATGTAATGTTATTAATTAACA 1380

Qy 4688 aacaagaagaataaagtcgtggaagaacccaacccactcgaatgaatgaacaaatgagc 4747

Db 1381 AACCAAGAAGTGAATATGTTGTTGAAATCCCAATCCACTTGTAAACGAATAATAGTGTC 1440

Qy 4748 ggtgtgagccgcatgctcaatagccagcgagaaagcggtctcaaggaagaataatca 4807

Db 1441 GATGTGATGCGATGCGCAAAATGACGAAGAAGATTCAGGTAGACAGGAAGAAATCA 1500

Qy 4808 catcgaggtgtactcaagcccgacccatcactccttcgaagcgagttttgtcccaagct 4867

Db 1501 CATGTGATGATGACTAAACCTGATCTTATCCACTTTTCGATGATTTTTCGCAAGTTCT 1560

Qy 4868 ctacttccttgagcatcctcctcctgctgacccatcgaatgltcctgacagctcatct 4927

Db 1561 CTACTCTCTAGAAATATCATTTTATTAATACATGTTAATATTAACAGTTTCATTT 1620

Qy 4928 aatagatcgatg 4939

Db 1621 AAAAAATGTAGC 1632

RESULT 22

PFAMSP834 1636 bp DNA linear INV 03-FEB-1999

LOCUS PFAMSP834

DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

ACCESSION D13361

VERSION D13361.1 GI:391800

KEYWORDS BEEF-like domains; major merozoite surface protein precursor.

SOURCE Plasmodium falciparum (isolate 834) merozoite, DNA.

ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 1636)

AUTHORS Jongwutives,S.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology, 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)

2 (bases 1 to 1636)

REFERENCE Jongwutives,S., Tanabe,K. and Kanbara,H.

TITLE Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates

JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)

MEDLINE 93295445

COMMENT Submitted (05-OCT-1992) to DDBJ by: Somchai Jongwutives

Department of Protozoology

Institute of Tropical Medicine

Nagasaki University

12-4 Sakamoto-machi

Nagasaki 852

Japan

Phone: 0958-47-2111x3747

Fax: 0958-47-6607.

FEATURES

source

1..1636

location/Qualifiers

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/dev_stage="merozoite"

1..1632

/gene="MSP1"

gene

CDS

<1..1622

/gene="MSP1"

/codon_start=3

/product="major merozoite surface protein precursor"

/protein_id="BAA02622.1"

/db_xref="GI:391801"

/translation="PLKTKSEESIQEDNYASLENEFVLSLEGLKLNLEKKKLS YLSSGHLIAELKEVKNKNTGNSSENTPVNNALSEYKKFLPEGTAVVAVSES GSOTLDSQPKRASHVIGASMTITTSOVDDVDVITVPIFGSEEDVDLGOV TGEAVTSPVIDNLISKTENEYEVLYLEPLGAVRSLKLEKNVMTFNVAKOILNSR FNKRENFNVLESDLIPKDLTSNVVKKPYKFLNKEKDKFLSYNTKDSIDVDI NFANDVLYGKYLSEKYSDDLISIKKYNKQENENEYLPFLNIEFLYTVNDKIDL FVHLEKAVNTYTESKNVEVKIKELMYLKTIDDKLADFKNNNPFVGLADSTDYNN NLTJFLSTGAVFENLAKTVLSNLDLGNLQGMNITSGHOCVKRCKPQNSGCFPHLDER EEKCLINTKQESKCYENNPPTCNENKGCDDADKCTEDSDSNKGIKTECTKDS YPLFDGIFCSSNMFGLISFLIIMLILYSTI"

BASE COUNT 695 a 199 c 239 g 503 t

ORIGIN Chromosome 9.

Query Match 18.4%; Score 907.2; DB 3; Length 1636;

Best Local Similarity 72.2%; Pred. No. 2.9e-189;

Matches 1179; Conservatly 0; Mismatches 453; Indels 0; Gaps 0;

Qy 3308 ctctcttgaagactctctccgagagagacacgagcaggaactaactacgagcctcg 3367

Db 1 CTCCATTAAACCTTTAAGTGAAGAAATCAATTCAACGAAAGATTAATATCCAGTTTAA 60

Qy 3368 agaactcaaggtcgtctgaagtcgaagcgaagtcgaagcgaagtcgaagtcgaagtcga 3427

Db 61 AAACTTTAAAGTATTAAAGTAAATTAAGTAAAGTAAAGTAAAGTAAATTAATTAAGTAA 120

Qy 3428 agaagaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcga 3487

Db 121 AGAAAAATTTATCTATCTATCAAGTGAATCAATTTAATTTGCTGAATTAATAAGAAAG 180

Qy 3488 tcattaaagaagaagaactacacccgcaatagcccaagcgagaataatacagaagtcgata 3547

Db 181 TAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 240

Qy 3548 aagcagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcga 3607

Db 241 ATGCATTAGATCTTACCAAAAAATTTCCACAGAAAGAAAGATGTTGCAACAGTTGTAA 300

Qy 3608 ctgaatcgtgctccgaagaacactggaagtcgaagtcgaagtcgaagtcgaagtcgaagtcga 3667

Db 301 GTGAAGTGGATCGACACATTAGAACAAGTCAACCAAAACCAACGACATCAACTCATG 360

Qy 3668 tcgagacggaatcccaatacaataccacatctcgaagtcgaagtcgaagtcgaagtcgaagtcga 3727

Db 361 TAGGACGAGACTCTAACACATTAACACATTAACACATTAACACATTAACACATTAACACAT 420

Qy 3728 tcatcatgtgctatctcctcgagagcgagagtcgaagtcgaagtcgaagtcgaagtcgaagtcga 3787

Db 421 TAAATCATAGTACTATATTTTGGAGAAATCCGAAGAAATTTATGATGATTAGAGCAATAG 480

Qy 3788 tcaccggtgagcgctgctactcctcctcgatgattgataactctgtccaataatcgagaagc 3847

Db 481 TAACAGGAGAAAGCAGTAACCTTCCTCCGTAATTAACATTAATTTTAATAATTAATAATG 540

Qy 3848 aatacgaagtcgtcatctgaagcctcgtgagcgctgataagtcgtcctaagaacagc 3907

Db 541 AATATGAGGTTTATATTTAAACCTTTAGACAGTGTTTATAGAAAGTTTAAAAAACAAT 600

Qy 3908 tggagaatacgtgagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcgaagtcga 3967

Db 601 TAGAAAAATTAACGTTATGACATTTAATGTTAATGTTAAGGATATTTTAAATTCACAGATTYA 660

Qy 3968 ataagagaagaatctcaagagcgtcttgagagcgagtcgattccctataaagacttga 4027

Db 661 ATTAACGTAAGAAATTTTAAAGATTTTAAATGATTTTAAATTTTAAATTTTAAATTTTAA 720

Qy 4028 cctccttaactacgtgttcaagagccatacaagtlcctcaataaagaagaaggagata 4087

```

Db      721 CATCAAGTAATATGTTGTCAAAGATCCATATTAATTTCTTAATAAGAAAAAGAGATA 780
Oy      4088 aatttcgtcctgttaccaatatataaaggacctccctgcaccggtatcaatttcgcta 4147
Db      781 AATTCTTAAGCGATTAATTAATTAATTAAGGATTAATTAATTAATTAATTAATTAATTA 840
Oy      4148 atgatcgtcgtggtatatacaagatccctcagcgaaataatacaagtcgactgactgcta 4207
Db      841 ATGATGTTCTTGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900
Oy      4208 ttaaaagtatatcaacgataagcgaagcgagaaatgaataatactgccttcctgtaata 4267
Db      901 TTAATAAAATATATCAACGACAAACAAAGGTGAATAATGAGAAATACCTTCCTTTTAAACA 960
Oy      4268 acatcgaaacctgtaacagcagtgtaacgacaaatcgaccttcgtaattccacctg 4327
Db      961 ATATTGAGACCTTATATTAATAACGTTAATGATTAATTAATTAATTAATTAATTAATTAAT 1020
Oy      4328 aggcacaagtcctcaactaactaactaagcagcaatgltggaagttaaatcaagagac 4387
Db      1021 AAGCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080
Oy      4388 tgaactactcaaaaacaaacccaagacagtcgcaatttcaagaataataaataattcg 4447
Db      1081 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140
Oy      4448 tcggaattcgacacgcgtctaccgataactaacaacacacacacacacacacacacacacac 4507
Db      1141 TTGGATTTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
Oy      4508 ccactgacatggtgtcgcgaacacctcgcaaacacagtcgtgacnaatctgctcgacgca 4567
Db      1201 GTACAGGATGTTGTTTGAATAATCTTGTAATAACCGTTTATCTAATTAATTAATTAATTAAT 1260
Oy      4568 acctgagggcctgctgcaacatctccagcaccaatgctgtaagaacagtcgccccaga 4627
Db      1261 ACTTGCAAGGATGTTTAACCATTTTCAACAACCAATGCGTAATAAACAATGTCACAA 1320
Oy      4628 atagcgcgctgttcagcactcgcgacgacgacgacgacgacgacgacgacgacgacgacgac 4687
Db      1321 ATTCTGATGTTTCAACATTTTGAATGAAGAAGACAAATTAATTAATTAATTAATTAATTAATTA 1380
Oy      4688 aacaagaagagataaagtcgltggaagaccccaaacctcactcgtcaatgaataacatgagc 4747
Db      1381 AACCAAGAGTATGATTAATGCTGTAATAATCCAAATCTCTACTGTAAGAAATAATTAATTAAT 1440
Oy      4748 ggtgtgacgcgactgataatgacccgaggaagacgagcgtcctaacggaagaataatca 4807
Db      1441 GATGTGATGCAGATGCCAATATGATCCGAAGAAAGATTGAGGTAGCAACGGAATAAATCA 1500
Oy      4808 catgcaagctacaaagcccgacactcctacactcctcgcgacgacgacgacgacgacgacgac 4867
Db      1501 CATGTGAATGTAATAACCTGATTCATTCACACTTTTCGATGATTTTCTGCAATTCCT 1560
Oy      4868 ctaattctcgtggacatccctcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 4927
Db      1561 CTAACTCTTAGAATATCATCTCTATTAATCTCATGTTAATATATATACAGTTTCATTT 1620
Oy      4928 aatagatgatg 4939
Db      1621 AAAAAATGTAGG 1632

```

RESULT 23

PFAMSP844 1636 bp DNA linear INV 03-FEB-1999

DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

ACCESSION D13363.1 GI:391816

KEYWORDS EGF-like domains; major merozoite surface protein precursor. Plasmodium falciparum (isolate 844) merozoite, DNA.

ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 1636)
TITLE Jongwulwies, S.
JOURNAL Direct Submission
Submitted (05-Oct-1992) Somchai Jongwulwies, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)
2 (bases 1 to 1636)
REFERENCE Jongwulwies, S., Tanabe, K. and Kanbara, H.
AUTHORS Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
MEDLINE 93295445
COMMENT Submitted (05-Oct-1992) to DDBJ by:
Somchai Jongwulwies
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi
Nagasaki 852
Japan
Phone: 0958-47-2111x3747
Fax: 0958-47-6607.

FEATURES
source Location/Qualifiers
1..1636
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/dev_stage="merozoite"
1..1622
/gene="MSP1"
/gene="MSP1"
/codon_start=3
/product="major merozoite surface protein precursor"
/protein_id="BA02624.1"
/db_xref="GI:391817"

gene
CDS
1..1622
/translating="PLKTLSEESIOEDBNVASENFKVLSKLEGLKDNLEKKLS
YLSGLHLIAELKEVIRKNNTGSPVNNVDVNNALLESYKFLPEGTDAVAVSES
GSDTLQSPKRPASTHYGAESENITTSQNDVDEVDVITVPIGSESDYDDGOVY
TGEATTPSYIDNLSKIEYEVILTKPLAGYRRSLKQLENNWTFVNVKDDIISR
FNKRENNFNVEDSLIPYKDLTSSNVVKDPIYKLRKRDFFSVITKDSIDTDL
NFANDVLYYKILSEKYSKSDLSIKYINDKGENEKYLPFLNNITELYKYVNDIDL
FYIHEAVLNTYERKSNVEVYIKELNLTITQDLADFFKNNNVGJADISTVDNHN
NLTKFELSTGWFENLAKTVLSNLDGNLQGLMTISOHCYKOCPEKSGCFRHLDER
EECKILNVKQGGDKCENPNPTCENNGCGDADAKCTBEDSGSKKTIETCTPDS
YPLFDGIFCSSNFIIGISPLILMLILYSFI"

BASE COUNT 694 a 198 c 240 g 504 t
ORIGIN Chromosome 9.

Query Match 18.4%; Score 907.2; DB 3; Length 1636;
Best Local Similarity 72.2%; Pred. No. 2; 9e-189;
Matches 1179; Conservative 0; Mismatches 453; Indels 0; Gaps 0;

```

Oy      3308 ctccctcgaagactcctcgcgagagagacatccagacgagataactacgcacgctcg 3367
Db      1 ctccatttaaaacttttaagtgaaagatcaattcaaacacagaatattttgccaagttag 60
Oy      3368 agaacttcaaggtcgtcgtcgaagcgtgaagcgtgaagcaaacctgaaactgagaga 3427
Db      61 AAAACTTTAAAGTATTAAGTAAATTAATTAAGGAAATTAAGGATTAATTAATTTAGAAA 120
Oy      3428 agaagaagctcagctacccctcctcagcgactgatacactgctgcgcggtccaaggaag 3487
Db      121 AGAAAAAATTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 180
Oy      3488 tcatlaagaacaagaactacacgcgcaatagcccaagcagagataataacagacgtgtaata 3547
Db      181 TAATATAAATAAATAATTTATACAGGTAATTTCTCCAGGTAAATTAATTAACGATGTTAACA 240

```

OY	3548	acgcacggaactcttcaagaagaagtcctgcctcttaagaagaacatgctgcgcactctggtgt	3607
Db	241	ATGCATTGGAATCTTTACAAAAAATTTCTCCACAGAGGAACAGATGTTGGACAGTTGTA	300
OY	3608	ctgaatctcgatctccgacacacactggagcagctctcaaccttaagaagccctgcattactatg	3667
Db	301	GTAAGAGTGGATCGCACACATTTAGAACAAAGTCACCAAGAAACAGCATCACTCATG	360
OY	3668	tcggagccggatcccaatacaatctcaagaagtcgaagatggtgcattgacg	3722
Db	361	TAGGAGCAGACTCTAACACATATAACAACATCACAAAATGTCGATGTGAAAGTAGATGACG	420
OY	3728	lcatcatgtgcccattctcggcgagagcggaagagactcagatgaacctcgccgaagtg	3787
Db	421	TAACTCATAGTACCTATATTTTGGAGAAATCCGAAACAACTTATGATGATTTAGCAACAATG	480
OY	3788	tcaccggtgagctgctcaactccttccegtgtgataacatctgtctcaaaatcgsagaag	3847
Db	481	TAAACAGGAGAGCAGTAAACCTCTCCGTAATTTATTAACATCTTCTAAAAATTTGAATAAG	540
OY	3848	aatccgaagtgctctatctcgaaacctctcgcaagcgctctatagtgctctccaagaagcgc	3907
Db	541	AATATGAGGTTTATATTTTAAAACTTTAGCAGGCTGTTTATAGAACTTTAAAAAACAAT	600
OY	3908	tgagaaataacgltgatgaccttcaatgtcaacgltgaagacatctcgaacagccgcttta	3967
Db	601	TAGAAAAATAAAGTTATGACATTTTAATGTTAATGTTAAGATATTTTAAATTCACAGATTTA	660
OY	3968	ataagagagaaaatttcaagaagcctcttggaagcgagcttgattccctctaaagccgta	4027
Db	661	ATTAACCGTAAATTTTCAAAAATGTTTAAATCAATCAATTTAATTCATTTAAAGATTTA	720
OY	4028	ccctccttaactgaactgltgtaagaagcccaatacagttcttcaataaagaaagagagta	4087
Db	721	CATCAAGTAATTTATGTTGTCACAAAGATCCATATAATTTCTTATATAAGAAAAAGAGATA	780
OY	4088	aattctgctagtctacaactatatacaaggtcccatcgacacgatatcaattccgcta	4147
Db	781	AATCTTTAAGCAGTTAATAATTTATTAAGGATTCATTAACATGTCGGATTAATAATTTTGAA	840
OY	4148	atgaigtgctgggtgatctacaagaatccgtgacgaaatatcaagttcgaccttgactcta	4207
Db	841	ATGATGTTCTTGGAATATTATTAATAAATNTATTCGCAAAAAATRTAATACGATTTAGTTCA	900
OY	4208	ttaaaagtatatacaacgataaagaaagcgagaaatgaanaatatctgcctctctgtaata	4267
Db	901	TTTAAAAATATATATCAACGACAAACAAAGGTGAATATGAAATACCTCCCTTTTAAACA	960
OY	4268	acatcgaaaccccttacaagaacgttgaacgaaatgaacatcgacctcttcgtaattacccgg	4327
Db	961	ATATTTGAGACCTTATATTAANAACGTTTATATGATATAAATTTGATTTATTTGTAATTCATTTAG	1020
OY	4328	agcgcaaggtctcctaactactactcagaagaagcaatltggaagfltaaatcaagaagc	4387
Db	1021	AAGCAAAAGTCTTAATTTATATACATATGAGAAATCAACAAGTAGAAGTTAAATTAANAAGAC	1080
OY	4388	tgaactacttcaaaaacaatccaagaacagctgtagatcttcaagaataatacaattcgg	4447
Db	1081	TTTAATTTACTTAAAAACAATTCACAGACAAATTTGGCAATTTTAAAAAATAATCAATTTTCG	1140
OY	4448	tcggaattgacagacctcttcaacgatataaacaacaacatctcctgaccaagttcgt	4507
Db	1141	TTTGAATTTGGGATTTTATTCACAGATTTATTAACATATAATTAACATTATTTAGCAAAAGTCTCTTA	1200
OY	4508	ccactgcatggtgtctgnaaacctcgccaaacagltgctgagcaatctgtctcgacggca	4567
Db	1201	GTAACAGGTATGCTTTTGGAAAATCTTGCTAAAAACGCTTTTATCTATTAATTAATTAATTAATTA	1260
OY	4568	accgcgagcgcatgcttgaataatctccagacaacaatgctggaagaagaacgttgcgcccgga	4627
Db	1261	ACTTTGCAAGGTATGTTAAACATTTTCACAAACCAACATGCTTAAAAAACAATGTCCAGAAA	1320
OY	4628	atagcgcctglttcaagcatctggaacgagcgcgaaagtgacaaagtctctcgtgaactaca	4687

Db	1321	ATTCTGGATGTTTCACACACTTTAGATGAAGAGAGCAATGTAATCTTTATTAAATTACA	1380
Qy	4688	aacaagaagagataaagtcgtcgtgagaaaccacaacctcctcgtcaatgaaaaaactctgcg	4747
Db	1381	AACAAGAAAGTGATAAATGTGTTGAAAAATCCAAATCTACTGTGAACGAAAAATTAATGCG	1440
Qy	4748	gctctgacgcgatctctaataatgcacccggaagaacagcgcccttaacgaaagaataata	4807
Db	1441	GATGCGATCCAAATGCGCAATGTACCGAAGAAAGATTACAGGTAGCAAGGAAAAATTA	1500
Qy	4808	catcgagctgaactaagcccgagctccctacacactcttcgaggatcttttgcaccgct	4867
Db	1501	CATGTCAAATGCTACTAAACCTGATCTCTTATCCACTTTTGCATGATATTTCTCGACATCTCT	1560
Qy	4868	ctaattccctcggacatccctccctcctcgtcatcccaatgctgaatcctcgtacagctcatc	4927
Db	1561	CTAACTCTTACGAATATCATCTTATTAATTAATCACTCATGTAAATATTAACATGTTCAATTT	1620
Qy	4928	aatagatcgatg 4939	
Db	1621	AAAAAATGTAGG 1632	

RESULT	24			
PfPRMSF8222				
LOCUS				
DEFINITION	p.falciptarum DNA for the precursor to the major merozoite surface proteins, C-terminal.	1636 bp	DNA	INV 03-FEB-1996
ACCESSION				
VERSION	D13359.1			
KEYWORDS	EGF-like domains; major merozoite surface protein precursor. Plasmodium falciparum (isolate 822/2 and 827) merozoite, DNA.			
SOURCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
REFERENCE	1 (bases 1 to 1636)			
AUTHORS	Jongwutiwes,S.			
TITLE	Direct Submission			
JOURNAL	Submitted (05-OCT-1992) Somchai Jongwutiwes, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)			
REFERENCE	2 (bases 1 to 1636)			
AUTHORS	Jongwutiwes,S., Tanabe,K. and Kanbara,H.			
TITLE	Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciptarum from field isolates			
JOURNAL	Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)			
COMMENT	93295445 Submitted (05-OCT-1992) to DDBJ by: Somchai Jongwutiwes Department of Protozoology Institute of Tropical Medicine Nagasaki University 12-4 Sakamoto-machi Nagasaki 852 Japan Phone: 0958-47-2111x3747 Fax: 0958-47-6607. Location/Qualifiers 1..1636 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /dev_stage="merozoite" 1..1622 /gene="MSP1" <1..1622 /gene="mSPI" /codon_start=3 /product="major merozoite surface protein precursor" /protein_id="BAO2620.1" /db_xref="gi:39179" translation="PLKTISEESIOTEDNVASLENFVLSKIEGKLNDNLNEKKKS"			
FEATURES				
source				
gene				
CDS				

YIISGLHHLIAELKIVINKNTYNSPSVNTVDNNALESIKKELPECTDVATVSES
GSDILEQSQPKRPASTHVGAESNTITTSQNDVEDVDVIVPESESEDYDLQVY
TGEATVPVINDILISKIENEYVELKPLAGVYRGLKOLENNWTFPNVNDILNSR
FNKRNKNVLESDILPYKDLTSSNVYKPKEDRDLFSSYNIKRSIDTID
NFANDVILGYIKLISEKYSDDLSIKKYINDOGEMEKLPENLNETYKTYNDKIDL
FYIHEAVLANTYERKSNEVYKIKELNLTKTQDLAPFKKNNVGLADLSTVDNH
NLITFELSTGVFENLANTVSNLJDGNQGLNTSQHOCVAKQCPQNSGCRHDER
BECKCLFNKOGSKVENPNPTENNNGGCDADKCKTEEDSGSKGKITCCTKPDPS
YPLDFGICSSSNFLGISFLITLIMLILYSFI*

BASE COUNT 694 a 199 c 239 g 504 t
ORIGIN Chromosome 9.

Query Match 18.3%; Score 905.6; DB 3; Length 1636;
Best Local Similarity 72.2%; Pred. No. 6.5e-189;
Matches 1178; Conservative 0; Mismatches 454; Indels 0; Gaps 0;

QY 3308 cccctctgaagactctctccgagagagacatccagacgagagataactcgcacgctcg 3367
DB 1 CTCGATTAAAACTTTAAGTAAATTCGAAAGATCAATCAACAGAGATTAATTCGCACTTTAG 60
QY 3368 agaacctcaagctcctcgtctaaagctcaagcgaagctgaagcaaacctgaacctgaga 3427
DB 61 AAAACCTTTAAAGTATTAACTAAATTTAGAAAGAAATTTAAAGATTAATTTAAATTTGAAA 120
QY 3428 agaagaagctcagctcctcctcctcagcgaagctcagctcagctcgcgagctcgaagag 3487
DB 121 AGAAAAATTTATCATACCTTATCATCAAGTGAGATTACATCATTTAATTCGTGATTTAAAGAA 180
QY 3488 tcaatbaagaacagaactacacccggaatagcccaagcgaagataatacagaagctgaata 3547
DB 181 TATATATAAAATTAATAATTAATACAGTAATTCACAGTAATTCGATTAATTAATTAATTAATTA 240
QY 3548 acgacactgaatctctcaagaagctcctcctcctgaagaacagaatcgtccgaactgctgt 3607
DB 241 ATGCATTAAATCTTTCACAAAAATTTCTCCCAAGAGAACAGATGTTTGCACAGTTGTAA 300
QY 3608 ctgaactcgtcctcgaacacactcgaagcagctcgaactgaagaagcctcgaactcgaatg 3667
DB 301 GTGAAAGTGGATCCGACACATTAGAACAAAGTCACCAAGAAACACGATCAATCAATCATG 360
QY 3668 tggagcgcgaatccaaataaataatccaaatcgaagctcgaagctcgaagctcgaagc 3727
DB 361 TGGAGACCAAGTCTTAACAAATTAACAAATCAACAAATGTCGATGATGAAGTATGATGAGC 420
QY 3728 tcaatcctcgtcctcctcctcgcgagagcgaagcgaactcgaactcgcgcgcgcgcgcgcgc 3787
DB 421 TATCATTAAGTACTTATTTGGAGATTCGAGAGAGAGATTAATGATGATTAAGGACAGTAG 480
QY 3788 tcaacgctgaagcgtcgaactcctcctcgcgtaactgaatacatctgtccaaatcgaagacg 3847
DB 481 TAAACAGGAGAACAGTAACTCCCTCCGTAAATGATTAACATTAATTTTAAATTTGAAAGT 540
QY 3848 aatacgaagctgctctcgaagaacctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3907
DB 541 AATATGAGGTTTATATTTTAAACCTTTAGCAGGTTTATGAAAGTTTAAATAAAACAAAT 600
QY 3908 tggagaataacgctgaactccaatgctcaacgctgaagagacatctcgaacagcgcgcctta 3967
DB 601 TGAATAAATATGACATTAAATGTTAATGTTAAGATTAATTTAAATTTTAAATTTTCAACGATT 660
QY 3968 ataagaagagaaactcgaagacgctcgcgagagcgaactgattcctcctataaagacctga 4027
DB 661 ATTAACGTCGAAATTTCAAAATGTTTGAATCAATGATTAATTTTCAATTAATTAAGATT 720
QY 4028 cctcctcctacacgctcgcgaagcgaactcgaagctcctcctcctataaagaagagagagata 4087
DB 721 CATCAAGTATATTTGTTGTCAAAGATCCATATAATTTCTTAATTAAGAAAAAGAGATA 780
QY 4088 aattctcgtcctgaactacatatcaagaagcctcgaacagcgaactcaaatctcgata 4147
DB 781 AATTTCTTAAGCAGTTTATTAATTAATTAAGATTCAATAGATACGATATAATTTTGGCAA 840

QY 4148 atgatgctggggatctcaagaatctcgaagcgaagaaatacaaatctcgaactcgaactccta 4207
DB 841 ATGATGTTCTTGGATTTATTAATAATTTATTCGAAAAATTAATCAATCGATTTAGATTCAA 900
QY 4208 ttaaaagratctcaacgataagcgaagcgaagataaataatactcgcctccctcgaata 4267
DB 901 TTAATAAATATATCAACGACAAACAAAGTGAAATATGAAATACCTCTCCCTTTTAAACA 960
QY 4268 acatcgaacccctcgaagcagltgaagcagaacacgcgaactcctcgaactcgaactcga 4327
DB 961 ATATTGAGACCTTATTAATAACAGTTAATGATTAATAATTTATTTATTTGTAATTCATTAG 1020
QY 4328 aggcgaagctcctcaactaactcgaagaagcgaatgctgaagcttaaatcaagagac 4387
DB 1021 AAGCAAAAGTTCTTAATTTATCATATGACAAATCAACCTGACGTTAAATTAATAAGAAC 1080
QY 4388 tgaactaactcaaaaatcgaagaacagctgcgaattcgaagaataaatacaattcgc 4447
DB 1081 TTAATTAATTAATAACAAATTAAGACAAATTTGGCAATTTTAAATAAATTAACAAATTTG 1140
QY 4448 tgcgaatcgaacccctcgtctacagatataacacacacatcctcgaagctcgt 4507
DB 1141 TTGGAATTCCTGATTTATCAACAGATTATACCATATTAATTAATTTGCAAAAGTTCTTA 1200
QY 4508 ccactgcatggtgctcgaacacccctcgcgaacagctcgtcgaagcgaactcgcgcgcga 4567
DB 1201 GTACAGGTATGTTTGTGAAATCTGTAAACCGTTTATCTAATTAATTAATTAATTAATTA 1260
QY 4568 acctcgaagcagctcgtcgaacatctcccaacacacagctcgaagaacagctcccccaga 4627
DB 1261 ACTTGAAGGTATGTTTAAATTTTCAACACCAACCAATGCTTAATAAATTAATTAATTAATTA 1320
QY 4628 atagcgcctcgtcgaagcactcgaagcgaagcgaagcgaagctcctcgaactcga 4687
DB 1321 ATTTCTGATGTTTACACACATTTATAGATGAAGAAGAAATGTAATTTATTAATAATTAATTA 1380
QY 4688 aacaagaagagataagctcgtcgaagcgaaccccaacccctcgtcgaactcgaagcga 4747
DB 1381 AACCAAAAGTATGTAATTTGTTGAAATTCACAAATCTCTGTAACGAAATTAATTAATTAATTA 1440
QY 4748 gctgtagcgcgaatgctcgaatcgaacgaagcgaagcgaagcgaagcgaagcgaagcga 4807
DB 1441 GATG 1500
QY 4808 catcgcagctgaactcgaagcgaactcgaactcgaactcgaagcgaactcgaactcgaact 4867
DB 1501 CATGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1560
QY 4868 ctaattcctcgtcgaactcctcctcgtcgaactcgaactcgaactcgaactcgaactcga 4927
DB 1561 CTAACTCTTAGGAATATATCTTATTAATCTCATGTTAATTAATTAATTAATTAATTAATTA 1620
QY 4928 aatagatcagtg 4939
DB 1621 AAAAAATGTAAG 1632

RESULT 25
PEMSMSAL 3395 bp DNA linear INV 29-JUL-1994
LOCUS P. falciparum gene for major surface antigen (MSA-1).
DEFINITION X61930 S37357
ACCESSION X61930.1 GI:9917
VERSION X61930.1 GI:9917
KEYWORDS major surface antigen; MSA1 gene.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 3395)
AUTHORS Certa, V.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-1991) U. Certa, F. Hoffmann-La Roche, Dept. PRB,
Grenzacherstr. 124, CH-4002 Basel, SWITZERLAND
REFERENCE 2 (bases 1 to 3395)

QY	1462	actaaagagtatgaaagctgcgtcgaaacgagatctatgatgccaaatccaataacatc	1521
Db	1636	AAAAAAGATTTATGAAAGATTACTTGA AAAAATTTTATGA AATGAAAATTTAAATAATATTTT	1695
QY	1522	gaccgcgaacctccggaagaatgagggaaacggtactcttaacaagtggaagtaactg	1581
Db	1696	GACCAAGATGCGTATGATATTAATTTTCAGTGCACAGATATATCATATATATGTTGAAAACAA	1755
QY	1582	acaacacaaatacccttgcataccctabgaaatctcaagcaatactctggaagctacc	1641
Db	1756	ATATATATATATATATATTTTATCTCTCTCAATATATTTCTGATATATATATGTTCAAAAATTTAAA	1815
QY	1642	aaagctcttaagtaatgatggagactatctctcggaacaacttgctggaagaanaacta	1701
Db	1816	AAGGCTCTTTATATCTTTGGAAGATTATCTTTTAAGAAAAAGAAATTTCTGAAAAACATTTT	1875
QY	1702	aagtaatacaagaaatctcataagtaagatcgaaacagagatcgagagctgttgagac	1761
Db	1876	AATCTATTTATTTACTTTGAAAACCTGGCCCTGAGCTGATATATATATTAACATGACAGAA	1935
QY	1762	attaagaagagtggaagaacgtgtgttggaagaagaatctacaagaagaacaaataaaca	1821
Db	1936	ATTAAGATAGTGAAGAACAAATTTCTATGAAAAAATTTTAAAGACTTAACATTCACACA	1995
QY	1822	gatggaagatcccggaagctcccgatattgtttaagtcgaagtcgaaggtgtccctc	1881
Db	1996	AATG --- CTCTCTTAGAAGATATCGATATTTGTAAATTTCAAGTACAAAAAGTTTATTA	2052
QY	1882	atgaaacaagatltgatgaactcaagaagactccaactatctcgaagaagltgagltaaa	1941
Db	2053	ATTTAAAAAATGAGAGCTTTAAAGAAATGTGAATTTATTTTAAAAATGACACACTATAAA	2112
QY	1942	cataatacatalgtccggaatagttatgaacgagagaatgaagaacacatacctc	2001
Db	2113	GATGATATTCTATGATCCAAATTTATTAACACAAATATTAACACGAGAACATATATTTA	2172
QY	2002	atcgtactcaagaagaatagatagacaactgaaagtgtcatgcccgaagtgcgaagctg	2061
Db	2173	ATTGATTTAAAAAAGAAAGTATGATTAATTTAAAGAAATTTATACCAAGTAAAGACATG	2232
QY	2062	atcaacgaagaagaagaacatataaacctggaagcagtgcaagataaactccgaagctcc	2121
Db	2233	TTAAAGAA -- AGAACAGAGCTGCTTTATCAAGTATTTACACACCTTTAGTTGCGAGACAGC	2230
QY	2122	acaagaagaagatataacccggaagctctacacccaaccccggaacaacagccggttcagct	2181
Db	2291	AAACACACTGAGAGATGGG - GGTCACTCCACACACACTTTATCCCAATCAAGAAACACAA	2349
QY	2182	ctcgaaagcgtatagcgtgtgcaagctcaagctcaagaagcagaagcagaagctccagtg	2241
Db	2350	GTAAACAGAAAGAAACGATAGGACACACAAACAGGTAACATTAATTTACACCAACACAA	2409
QY	2242	ccagtgcccgcttccgaaggtctaaagctcaagtgctctacacccacccgctctgtgatatac	2301
Db	2410	CCATCACACCAAAAAGAAAGTAAAGTTGTTGAAAATTTCAATATGAAACATTAAGATTAATAC	2469
QY	2302	aagaccggaatltcagcaaacctgagactactcttgagaagctctatagagttccgtgaaca	2361
Db	2470	AATTCACAGAGCTTGACAAAAAAGATTTATCTTAACAAAAATTTAGATGAATTTTAACTATAA	2529
QY	2362	tccctacatctcgcaaaatatactctgtctctccaaagcaactgaagagaagagttctt	2421
Db	2530	TCATATATATGTCATTAATATATTTTATGATCAAACTATGATATGACCAAAAAATTTATTA	2589
QY	2422	aaacagatacaagatacacaagaagaagagatgaactgtccctctgtatccatctgagac	2481
Db	2590	GAGGATATATATCTTACTCCAGAGAAAGAAATGAATTAATAATCAATGATCATTAATTAAT	2649
QY	2482	ctgctgttcaatatccagaacaacatcccggtatgtatctatgtctcgtatagctccaac	2541
Db	2650	TTATATTTTATATATTTCAAAAAATACATACCTGATGATATTCATTTATATGATAGTATGAC	2709

QY	2542	aattctctctccacgttctcatgtagtataatgagagagatgcttgcgaactgcat	2601
Db	2710	AATGATTTACAAACATCTCTTTTGTGAATATATCAAAAGGAAGAATATTATTTACAT	2769
QY	2602	aaactcaagacaacgcacaagattaagaacctctcgagga	2642
Db	2770	AAACTAAAGAGAGAAATACATCAAAAAATATATAGAGA	2810
RESULT 26			
PPAP190A		3518 bp	linear
LOCUS			INV 26-Apr-1993
DEFINITION		P.falciiparum merozoite 190 kD precursor protein (p190) mRNA, 5'	
ACCESSION		end.	
VERSION		M35727	
KEYWORDS		M35727.1 GI:160549	
SOURCE		p190 gene; variable surface antigen.	
ORGANISM		Plasmodium falciiparum (individual isolate RO-33 Ghana) cDNA to mRNA.	
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS		1 (bases 1 to 3518)	
TITLE		Certa,U', Rotmann,D', Matile,H. and Reber-Jiste,R.	
JOURNAL		A naturally occurring gene encoding the major surface antigen	
MEDLINE		precursor p190 of Plasmodium falciiparum lacks tripeptide repeats	
FEATURES		EMBO J. 6, 4137-4142 (1987)	
source		88166657	
		Location/Qualifiers	
		1..3518	
		/organism="Plasmodium falciiparum"	
		/isolate="RO-33 Ghana"	
		/db_xref="taxon:5833"	
mRNA		1..>3518	
		/note="RO-33 mRNA"	
		/product="p190 protein"	
		55..>3518	
		/note="CAMP mRNA"	
		/product="p190 protein"	
mRNA		203..>3518	
		/note="MAD20 mRNA"	
		/product="p190 protein"	
mRNA		266..>3518	
		/note="K1 mRNA"	
		/product="p190 protein"	
		337..>3518	
		/codon_start=1	
		/product="p190 protein"	
		/protein_id="AA29715.1"	
		/db_xref="GI:160550"	
CDS		/translation="MKIIIFPISCFPIINOCVTHESYOELVKKALEDAVLTG	
		LFQEKWYKLGANTQYAKRADAVSFQSNACNPPGAYIPSTASTKALISRGANS	
		DDSSDSKSTADLHRVONTLFTEKLPTELFDLTNNHMLTDCNTHGRVTLIDGVE	
		EINBELKRLNFPDLRAKLADVCANDYCOQIPFMUKIRANLADVLRKLVGYRPLDFE	
		IKDWGKMEDEIKRKNKTIANINELIBSGSKTIDQNNADNEBKKLYQAQDFTFY	
		NKQEAHNLSVLEKRIDLTAKNNENIKLELDEDIKIRIDAEKTPGVNOLISLREK	
		ESRHEBKKEIAKTIKENIDRLPTDPLELEYLRKKKKVQVPEKSDPPKSVQIPKV	
		PYNGVLYPIPLDIHNSIADNDKNSGGLMNHTEKINIKETITDNKEKKIENNIN	
		KQIDLEKININHTREONKKLDEYBESKADYBEELKEFTYEMKNNPKDYVKTIS	
		ARVYVNEKQRYNNKPFSSSNSNVNOVLKALSYLEDYSLRKGISERDPNHYTLTNT	
		GLADIKRLTEIBKISSEKMLEKNFKGLTHSANSLEVSIDVLRQVQVLLIKITEDL	
		RKIEELFKLNAOLKDSIHVPNPIKPEYQNPPEYLLVLRKEVDLKEEFTPKVADMLKEDQ	
		AVLSITQPILEASSETTEDGSHHTLSQGETEETETETVGHVTTVYITLPPKRV	
		KVNSNTEIHRKSNDSOALTKTVYIKLIDELTKTSYICHKYLIVSNSSMDOKLLEVNL	
		TPRENEKSCDRDLRLNIRIONNIPAMSLIDSANNDQHLFPELYOKEMYYLHKLAK	
		ENHAKLLEERPQITVSTSSPGNTVYNTAQSATHSNQSQSNQSNASTYQNGVAVS	
		SGPVAVERSHDPLTVLSISNDLKGIVSLNAGNTKQVNPPLTISTEMEKREYNILKI	
		MLPIFNDDIKQFVKSNSKVIVGLTETQNALNDEIKRLKDTQLDSFDLVNKKYKTLKDR	
		LENKKLELGODKMQIKRKLTLKEQLESKINS"	
		1306..1347	
misc-feature		/note="This frameshift results in 11 new amino acids not	
		present in MAD-20 or K1; frameshift mutation FS-1"	
misc-difference		2496	
		/note="D1 in MAD-20 mRNA"	

misc_difference 2544 /note="D2 in MAD-20 mRNA"
misc_feature 3247..,3270 /note="results in a single point mutation shifting reading
frame to K1; frameshift mutation FS-2"
BASE COUNT 1570 a 430 c 437 g 1081 t
ORIGIN

Query Match 14.4%; Score 711.8; DB 3; Length 3518;
Best Local Similarity 56.8%; Pred. No. 3,9e-146;
Matches 1511; Conservative 0; Mismatches 1072; Indels 76; Gaps 8;

QY 10 atgaataatcttctccctcgtcattctcgttttcaatcaactacgtgctg 69
DB 337 ATGAAGATCATATCTTTTATGTCTTCTTTTATATTAATACCAATGTGTA 396
QY 70 acccaagatccatcagagagctggttaagaactggaagcttggagatgcctc 129
DB 397 ACAATGAAGTATCAAGAACTGTCAAAAACCTAGAGACTTGAAGATGACGATGTG 456
QY 130 accggaatacagcctgtcccaagaagaagatgctgtaataaggaagcagtcagc 189
DB 457 ACAAGTATAGTATTATTCAAAAGAAAATGATTAAGATGAAGCAAAATCTCAA 516
QY 190 gccgttaacaaccagcacaccgcgtctcaagagctgctgtaagcgtgagcgtgagc 249
DB 517 GTTGTTCAAAAGCTGCAAGATGCTGTAAGTACTAAAGTCTTAATAATCCTCCAGTCT 576
QY 250 tctgtgaccttgagggttccgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 309
DB 577 ACAGTACCTTCAGTACTGCAAGTACTAAAGTCTTAAGATCTCCAGTCTGCA--- 633
QY 310 gcaagcgcggttccgggaagcgtcgaagaaccatccatctgcaactctagcattcc 369
DB 634 -----AATCCTTCAGATGATTCAGAGATTC 660
QY 370 gaccccaagtcctacgc 429
DB 661 GATGCTAAATCTTACGCTGATTTAAACATAGAGTTCAAAATTCCTGTTCACTATTAA 720
QY 430 gaagctgaagtaaccacagctgttcgacctcaactacatagctgacacgtgtgataac 489
DB 721 GAAGCTAAATATCCGAACCTTTGATTTAACCAATCATATGTAATCTTGTGATTAAT 780
QY 490 attcagtgctcaaatatctgattgacggttaagagatcaatgcaactcgtatacag 549
DB 781 ATTCATGTTTCAAAATTTAATTGATGATATGAAACAAATTAATGATTAATTAATAA 840
QY 550 ttgaattctactcgcgtctgtaaggccaactgtaatacgttttcgcgaactgactat 609
DB 841 TTAACCTTTTATTTGATTTATTAAGACAAATAATTAATGATGATGCTTAATGATTAAT 900
QY 610 tgcataatctcatcaatttgaagatcagagccaacgagttgagcgtatgaaagattg 669
DB 901 TGTCAAATACCTTTCATCTTAATAATCGTCAAAATGAATAGACGACTTAATAAACTT 960
QY 670 gtcttcgatatcgcaagcctctcgacacatcaagaacatgtgggaagatggaagat 729
DB 961 GTGTTCGATATGAAACCATTAAGACTTTAATTAATGATATGAGCAAAATGGAAGAT 1020
QY 730 tatatttaaaagaataagacatcgagaacttaagacgtcgatcgagaatcccaaa 789
DB 1021 TACATTAATAAATAATTAACACCATAGCAAAATTAATGATTAATTAATTAATTAATTAAT 1080
QY 790 aagaccatagacaataaagaatgcaacaagaggaagaagaagaagattgatacag 849
DB 1081 AAAACATTTGATCAATTAAGATGCAAGATTAATGAAGAAAGAAAATTAATTAACAA 1140
QY 850 gcccaagtagacactgtcactatatacaaacagcttgaagaagccataactcactacgc 909
DB 1141 GCTCAATATGATCTTTTATTTACAAATTAACAAATTAACAAAGACATTAATTTAATGAC 1200

QY 910 gtactgagaagcgcatagacacccctcaagaagaatgaataatcaagaactgctgcac 969
DB 1201 GTTTTACAAAACGATTTGACACTTTAAAAAATGAAAAACATTAAGAAATTAATCTGAA 1260
QY 970 aagattatgaatlaagaatcct--ccgcgaagccaactctggaacacccctaacagc 1027
DB 1261 GATATGATTAATAATTAATAATGATGATGCGGAAAAACCCACTACTGAGTGAACCAATCTC 1320
QY 1028 tgcgtgacaagaacaagaatagagagcagcagaagaagatcaagaagatcgccaaa 1087
DB 1321 TCCCTGAGATTAGAAAAAGCTGAG--ACACGAAGAAAAATAAATAAATTAATGCGCAAA 1378
QY 1088 ccatttaagtcacatagatctccttactgtatcccttcgagctggaactactgtga 1147
DB 1379 CTATTAAATTTAATGATGATGATTTATTTACTGATCCACTTGATTAATGATTAATTTAA 1438
QY 1148 gagaagaagaataatagacatctcgcgccaagtgcga-----gacaaaggaat 1198
DB 1439 GAGAAAAAATAAATAAAGTTGATTAACACCTTAATCACAAGATCTACGAATCTGTTT 1498
QY 1199 caaccgaacctaatgaatatccatgtgtgagctacccctgtcttataacgatata 1258
DB 1499 AAATACCAAAAGTTCTTATCCAAATGATGATTTATTTACTTACCTCACTGATATTC 1558
QY 1259 acaa-----cgctcacaagcgtcaataagcttcggtgacttgatlaacccttcg 1309
DB 1559 ATATTTGATTAAGCTGCAAGATTAATGATTAATAATTCATATGCGATTAATGATTCCTATA 1618
QY 1310 attatacgaagaaccccttaagaatatctacacagaacatgagagaagaagatttca 1369
DB 1619 CTAAAGAAAAAATAATTAAGAAAAAATTTATTTACATTAATTAAGAAAAAATAATTAATTC 1678
QY 1370 acgaataatcagagagaagaatcaaatgtgaagaagaagaatgagagcagaagaagct 1429
DB 1679 ATTAACATTAATAAACAATTTGATTTAGAGAAAAAACAATTAATCAACAACAAGACAA 1738
QY 1430 acgaagccgcgagaagaagcttaacagatatacctaagaatgataaagctgctgacg 1489
DB 1739 ATTAATAATTAATTAAGATTAATGATAAAGTCAAAAGAAAGATTAAGAAATTAATTCGAA 1798
QY 1490 agatctatgattccaaatcaacaataacatcgacctgacccaactctgagaagaattg 1549
DB 1799 AATTTATGAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1858
QY 1550 gaaaagctactcttcaagaatgagaagaactgcaacccaataactcttgatcctatg 1609
DB 1859 GTGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1918
QY 1610 agaattctgaatcaatcttgagaagctcaaccaagctcttaagtatatgagactatt 1669
DB 1919 ATTAATCTGATTAATTAATGTTCAAAAATTAATAAAGCTCTTTTCAATCTTGAAATTAAT 1978
QY 1670 cctcgcggaacatgtgtgtgagaagaagaacttaagaatatacagaatctataaglaa 1729
DB 1979 CTTTAAGAAAAAGAAATTTGCAAAAAGATTTTATTAATTAATTAATTAATTAATTAATTAAT 2038
QY 1730 tcgaagaacgagatcgagagcgtgtgtgaagaacatlaagaagaagatgagaacagctgt 1789
DB 2039 TCGAAGCTGATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2098
QY 1790 agaagaagatttaacaaagaagaataacacagatgagaagatctcgtgagctccgata 1849
DB 2099 AAAAAAATTTTAAAGGCTTAACACATTCACGAATGCTTC---CTTGAAGTATCTGATA 2155
QY 1850 ttgttaagtcacaagtgcaagaagtgctccctcatgaaacagatgataaactcaagaaga 1909
DB 2156 TTGTAATAATTAACAAGTCAAAAAGTTTATTAATTAATAAATAATGAAGACTTAAGAAGA 2215
QY 1910 ctcaactcattctgaagaagctgtgagttaaacataatatacagtgtgcgaatagttata 1969
DB 2216 TAGAATTAATTTTAAATAAATAAAGCACTAAAGATTAATTAATTAATTAATTAATTAAT 2275
QY 1970 agcagagaaatagcaggaacataactactcactcgtactcagaagaagatagacaac 2029

```

Db 2276 AACCAAAATTAACCAAGACATATATTATTATTAATAAAGAGATGATMAAT 2335
Qy 2030 tgaagtgtctgctgcaagagcctgatacaagaagaagaacattaa 2089
Db 2336 TAAAGCAATTTATACCAAAAGTAAAGACATGTTAAAGAAAGACAGCTGCTTATCA 2395
Qy 2090 ctgaagagcagtcagataaactccagcctccacagaagaagatgaacccgagccta 2149
Db 2396 GATATACCAACCTTATGTTG-----CAGCAAGCAAAACACTGAAGATGGG 2443
Qy 2150 ccaccaagcccgagacaacagccggttcagctctgaaagcgatagcgtgcaagctcaag 2209
Db 2444 GTACCTCCACACACACATTTATCCCAATCAGAGAAACAAAGTAAACAGAAAGAGAA 2503
Qy 2210 ccaagagagaagcagcagcagcagcagctgcccagctgcccgttcacagagcctaagtc 2269
Db 2504 AAACAGTACAGACACACACACAGCTAACATATACCTACCACAAAGAAAGTAAAGCTTG 2563
Qy 2270 aagtgctcaccacacagcctcctgtaataacaagaacgagaatgacgaaactgact 2329
Db 2564 TTGAATTAATTCATTAAGACATTAAGATGACAAATTCACAAAGCCTTGACAAACAGTTT 2623
Qy 2330 acctgagaagctctatggttcctgaataacatccatctgcccacaataatctcgt 2389
Db 2624 ATCTTAAGAAATTAAGTGAATTTTAACTTAATCATATATATGTCATAATATATTTTAA 2683
Qy 2390 tctctacacgacactaagaagaaatcttaacagtaacaagaatacaagaagag 2449
Db 2684 TATCAAACTCTGATGATGAGCAAAATTTATAGAGTATATATATCTCTCCAGAGAA 2743
Qy 2450 aagaataaactgtcctctgtgatcagcagcctgtgtaataatccagaacaacatc 2509
Db 2744 A--TGAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2800
Qy 2510 cgttatgtatctatgtctgctgagctcaacaatctctctcctcagcgttcatagaga 2569
Db 2801 CTGCTATGATTCATATATGATAGTATGACATATGATTTTACAAACATCTCTTTTGAAT 2860
Qy 2570 tatatagaagaagatgctgcaacgttataaactcaagaacaagaagaatga 2629
Db 2861 TATATCAAAAGAAATGATTTATTTATTTACATTAAGAAAGAAAGTAAAGTAAAGTAA 2920
Qy 2630 acctctggaagaagctaa 2648
Db 2921 AATTATTAGAGAGCCAAA 2939

RESULT 27
LOCUS PEP190G 3518 bp DNA linear INV 29-NOV-1987
DEFINITION Plasmodium falciparum surface antigen p190 gene fragment.
ACCESSION Y00087
VERSION Y00087.1 GI:9924
KEYWORDS p190 gene; surface antigen.
SOURCE Malaria parasite P. falciparum.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 3518)
AUTHORS Certa, U.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-1987) Ulrich C., Hofmann-La-Roche, Department
2Ff, Grenzacher Str. 124, 4002 Basel, Switzerland
AUTHORS Certa, U., Rotmann, D., Matile, H., and Reber-Liske, R.
TITLE A naturally occurring gene encoding the major surface antigen
REFERENCE precursor p190 of plasmodium falciparum lacks tripeptide repeats
EMBO J. 6 (1987) In press
FEATURES
location/Qualifiers
1..3518
/organism="Plasmodium falciparum"
/strain="Ro-33"
/db_xref="taxon:5833"

```

```

CDS
/codon="Ro-33/1"
337..3518
/note="p190 surface precursor (AA 1-107) (3518 is 2nd base
in codon)"
/codon_start=1
/protein_id="CA68280.1"
/db_xref="GI:9925"
/db_xref="SWISS-PROT:P19598"
/translation="MKILFLCSLEFLEIFINTOCYHESYOELVKLEALDEAVLTGYS,
LFQEKMYLKGANTOVAAKPADAVASAKNPDPATVSTAGSKGIRSPGAMP,
DDSSDSDKSYADLKHRYONLYITKEIKYELPELJTNHMLIDCNHGFLLDGYE
EINELLYKINFPDLRAKLNDVANDYCOIIPNKIRANLNDVILKIKFEGYRPLDF
IKDNGKMEYIKKKRTTANINELIBESKTTIDONKANDBEKKIKYOYDLFTY
NKLOEAHNLISVLEKRIDTLTKENIKLLEIDDKLIDAEKPTGQNLISLEK
ESRHEEKIKEIKATIKFNIDRLFDPELELYREKKNKVDTPRSODPTKSVQIPK
PYNGIYVPLDLDHNSLADNDKNSGDIAMPNTEKEINEKITTDOKERKIFINI
KKOIDEKKNHNTKEONKLEDEKSKYBELLEKFEKFNENFNKDVOKIFS
ARYNYNVEKORYNKFSSNNSVYVOKTKKALSTLEPYSLEKIGISEDFHNYTIKT
GLEADIKLITEIKSEKTIIEKNKGLTHSNASLEYSYDLYKQVAVLYLKIETDL
RKIELFLKMAOLKDSIHVKNYKPKQNKPEPYLLVLEKVEKLEFIPKVDMLKEE
AVLSIITPLVAASETTEDEGSHSTHLSQSETEVEETEGHTTAVTITLPPKE
KVENSIIEKKSNDNSOALTKTYLKLDEFLEKSYCHRYLLVSNSSMDOKLELYN
TPEENELKSCDRIDLFNIONNIPAMYSLYOSMNDLQHLPELEKOKEMYIYHLKKE
ENHITKLEPEKQITGTSSTSPGNTVNTNOSATHSNOSONOSASTNOMCVAVS
SGPAYERSHDPLTYLSTISNDIKGLVSLNAGTKYVNPPLTISTBEKRYENIKI
MPLINDDIKOFVNSNRVITGLDETQNALNDELTKLTDLSFDLYNRYKKLDR
LENNKRELGDORMOIKRLTLEQLESLNS"
BASE COUNT 1570 a 430 c 437 g 1081 t
ORIGIN
Query Match 14.4% Score 711.8 DB 3; Length 3518;
Best Local Similarity 56.8% Pred. No. 3.9e-146;
Matches 1511; Conservative 0; Mismatches 1072; Indels 76; Gaps 8;
Qy 10 atgaataatcttctcctcctcgttcattcctgctttatcaatacactcagtcgtg 69
Db 337 ATGAATATATATCTCTTTTATGTTTCATTTCTTTTATTAATAATACAAATGTGA 396
Qy 70 accacgaatcctatcagaagcgtgltaaagaacttgaagacttgaagatgcgcctt 129
Db 397 ACACATGAAGAAGTATACAAAGACTTGCAAAAACCTGAGAGCTTTAGAGATGACGATATG 456
Qy 130 accgatacagcctgttccagaagaagaatggtcgtgaatgaagaagcagctgacg 189
Db 457 ACAGGTTATAGTTTATTTCAAAAGAAAGATGATTAAGATGAGCAATATACTCA 516
Qy 190 gccgttacacaagcagaccgcggttctaagggtctggtcagcgtgtgcgcggtg 249
Db 517 GTTGTGCAAAAGCCTGCAATGCTGTAAGTACTCAAAAGTGCTAAATCTCCAGTGCT 576
Qy 250 tctgtgacctgtggtgtcgttcgctcgcgcgcgcgcgcgtgcatcaagtggtcagt 309
Db 577 ACAGTACCTTCAGGTACTGCAAGTACTAAAGGTACTAAAGATCTCCAGTGCTGCA--- 633
Qy 310 gcaagcgcgcgttcgcgggaacgltcgaagaaccaatcactgcgaacactcagcattcc 369
Db 634 -----AACTCTTCAAGATGATTCAGAGTGA 660
Qy 370 gacgcgaagtcctcagcgcagcctcaagcgcagctgagaaactatctcactcaatcaag 429
Db 661 GATGCTAAATCTTACGCTGATTTAAACATAGAGTTCAAAATTTACTTGTCTCACTTAA 720
Qy 430 gacgtgaatcaccacagltgttcgacctcaactaatcatatcgtcagcagtgataga 489
Db 721 GAACATAATATCCGAACACTTGTGATTAACCAATCATATGTAACTTGTGTATAT 780
Qy 490 atcatgtctaaatatgtatgaaggttaacgaagaatcaatgaactcctgtacaag 549
Db 781 ATTCAATGTTTCAAAATATTAATTTGATGATGAAGAAATTAAGCAATTAATTAATAA 840
Qy 550 ttgaatttacttgactgtcgaagcacaactgaagaagttgagccaatgactat 609

```

Db 841 TTTAACTTTATTTGATTATTAAAGCAAAATTTAAATGATGTGCTTAATGATTAT 900
 Qy 610 tgcgaattccattccatttgaagatcagagcccaagattggaagctatagaagatlg 669
 Db 901 TGTCAAATFACCTTTCAATCTTAAATTCGTGCAATATGATGACGTACTTAAAAACTT 960
 Qy 670 gtcttcgatatcgaagcctctcgacaacatcgaagacaattggtgaagatlggaat 729
 Db 961 GTGTTCGATATGAAAAACCATTTAGACTTTATTAAGATATATGAGAAAAATGGAAGAT 1020
 Qy 730 tatattaaagaataagaagaccafcgagagaactaaccgagctgatacgaaatccaaa 789
 Db 1021 TACATTTAAAAAATTAACCAACCATATGCAATTAATGATTAATTAAGGAAGTAAG 1080
 Qy 790 aagaccctagacaaaataagaatgcaacaaagagagaagaagaagaagatgtgtaccag 849
 Db 1081 AAAACAATTTGATCAAAATTAAGAAATGCAATATGAAAGAAAAAATTTATACCAA 1140
 Qy 850 gcccaagtaagcagctgtccatctatacaaacagcttgaagaagccataccatcagc 909
 Db 1141 GCTCAATATGATCTTTTATTTTACAAATTAACAAATTAACAGAGACATATTTAATTAAGC 1200
 Qy 910 gtactggaagagcgcatagacacccctcaagaagaaatgaataatcagaagatgtcgac 969
 Db 1201 GTTTTACAAAAACGTATTTGACACTTTAAAAAATGAAAAACATTAAGAAATTTACTTGAA 1260
 Qy 970 aagattaatgaatlaagaatcctt--cgcgacgcaactctgggaacacccctcaacg 1027
 Db 1261 GATTTAGATTAATTAATTAATTAATGATGCCGAAAAACCCACTGCTGAGAACCAATCTTC 1320
 Qy 1028 tgcgtgcaagaagaagaagatagagagcagaagaagatcaagaagatcgcacaaa 1087
 Db 1321 TCCCTGAGATTAGAAAAAGAGTGCAG--ACACGAGAAAAAATTAAGAAATTTGCCAAA 1378
 Qy 1088 ccattgaagctcaacatgattctctcttactgatacccttgagctgagatgactacttga 1147
 Db 1379 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1438
 Qy 1148 gagagaagaataagaataatagacatctcgcgcaagtcga-----gacaaagaat 1198
 Db 1439 GAGAAAAAATTAATAAAAGTTGATGTAAACCTTAATTAATTAATTAATTAATTAATTAAT 1498
 Qy 1199 caacggaacctatgaatlatcccaatgtgtgacgtacccctgttcttataaagatac 1258
 Db 1499 AAATACCAAAAGTTCCTTATTCACAAATGTGTAATTCCTTTTCCATCCATCATATATTC 1558
 Qy 1259 acaa-----cgctctcaagagctcaatagatcctcggtgactgtatcaaccctcg 1309
 Db 1559 ATATATTCATTAGCTGCAGATTAATGATTAATAATTCATATGATTTAATGATCTCTATA 1618
 Qy 1310 attatacgaaagaacctcttaagaatatactacacagacaaatlgagaagaagaatgttca 1369
 Db 1619 CTAAAGAAAAAATTAATGAAAAAATTAATTAACAGTAATTAAGGAAGAAAAATATTCATTA 1678
 Qy 1370 acgaatacaagagaagaatcaaaatlgagaagaagaatltgagatgagacagaagaatgt 1429
 Db 1679 ATTAACATTTAAAAAACAATTTGATTAGAGAAAAAACAATTAATCAACAAAGAACAAA 1738
 Qy 1430 acgaagacgcgcagcaaaagctcaaaagatatacctaagaagatgaagaagctgtcgaag 1489
 Db 1739 ATTAATAAATTAATCTTTGAAAGATTATGAAAAAGTCAAAAAAGATTTAAGAAATTTACTTAAA 1798
 Qy 1490 agatctatgtatcacaattcaacaataacatcgacctgagcaactctgagaagaatgtg 1549
 Db 1799 AATTTTATGAAATGAAATTTAATTAATTAATTTTAACAAGATGTCGTGATTAATAATTAATTA 1858
 Qy 1550 gaaaacggttactcttacaagltgagaagaactgacacaccaataacccttgatcctatg 1609
 Db 1859 GTGCAAGATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1918
 Qy 1610 agaatcttaagcataactctgagaagctcacaagctcctaagcttcaagatacagagactatc 1669
 Db 1919 ATTAATCTGTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1978

Qy 1670 ctcgcgaaacattgtctgtgagaagaactaaagatactaaagaatctcataagtaaga 1729
 Db 1979 CTTTAATAAAAAAGAAATTTTCGAAAAAAGATTTTAATTCATTTATATCTTTGAAAACTGCC 2038
 Qy 1730 tgcgaagaagatcgaagcgttgtltgagaacatlaagaagaatgaagaagaactgtgttg 1789
 Db 2039 TCGAAGCTGATATTAATAAATTAATTAACAGAGAAATTAAGAGTAGGAAAAATCAATTCCTAG 2098
 Qy 1790 agaagaagatatacaaaagacgaanaataaccagaatgagaagaatccctgaggtctcgata 1849
 Db 2099 AAAAAATTTTAAAGGACTTAACACATTCACCAATGCTTC---CTTGAAGATATCTGATA 2155
 Qy 1850 ttcttaagtcacagatgcaaaaaggtgtcctctctatgaacagaatgtgaactcaagaaga 1909
 Db 2156 TTGTAATAATTAACAAGTACAAAAAGTTTATTAATTAATAAATAAGAAAGCTTTAAGAAAGA 2215
 Qy 1910 ctcaactcattctgaaagaagctgtgagttaaacataatatacatgtgtgcgaatagtata 1969
 Db 2216 TAGAATTAATTTTAAATAATGACCAACTAAAGATAGTATTTCAATGTCACAAATATTTATA 2275
 Qy 1970 agcagaagaataagcagaagaccatactactcactcgttactcaagaagaatagacaac 2029
 Db 2276 AACCACAAAATTAACCAAGAACCATATTTAATTTAATTAATAAATAAAGAAAGTATTAAT 2335
 Qy 2030 tgaagtggttcatgtcccaagctcgagagcctgtatcaacgagaagaagaagaacatlaaa 2089
 Db 2336 TAAAGAAATTTATTCACAAAGTAAAGACATGTTAAAGAAAGAACAGCTGTCTTATCAAA 2395
 Qy 2090 ctgagaagacagctcagataaactcagagctctccacagaagaagagataaccgcaagctta 2149
 Db 2396 GTATTACACACCTTTTATGTTG-----CAGCAAGCAAAACAATGAAAGATGGGG 2443
 Qy 2150 ccaccaagcccggaacacagcggttcagctctcgaagcgataggtgtcgaagctaaag 2209
 Db 2444 GTCACTCCACACACACATTTATCCCAATCAGAGAGAAACGAAAGTAACGAAAGAACAGAG 2503
 Qy 2210 cacaagaagcagaagcagcagacagctccaggtgcaggtgcccgtctcgaagctaaagctc 2269
 Db 2504 AAACAGTAGAGACACACACAGAGGTTAACATTAATTAATTAATTAATTAATTAATTAATTA 2563
 Qy 2270 aagtgcttaccacacacagctctcgtgtgaatacaagaagcagaatgtcgaagcaactgact 2329
 Db 2564 TTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2623
 Qy 2330 accttgaagcgtctatgaggttccctgaatacactcactcactcgcacaaatatactcgt 2389
 Db 2624 ATCTAAAGAAATTAAGATGAATTTTAACTTAATCATATTAATTAATTAATTAATTAATTA 2683
 Qy 2390 tctctcagacactatgaagagaagaatcttcaaacagtaacaagataccaagaagaagag 2449
 Db 2684 TATCAAACTCTAGTATGACCAAAAAATTAATTAAGATTAATTAATTAATTAATTAATTAATTA 2743
 Qy 2450 agagtaaacgtgtcctctgtgataccactggaagcgtgttccaatataccagaacaatc 2509
 Db 2744 A---TGAATTAATAATCATGATGATCGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2800
 Qy 2510 ccgttatagtatctatgttctgataagcctcaacaatctctctcactcactgttcaatgaga 2569
 Db 2801 CTGCTATGATTTTCAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2860
 Qy 2570 tataatgagaagagatgtgtgcaacctgtataaacccaagaagaagaagaatgaaga 2629
 Db 2861 TATATCAAAAAGGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2680
 Qy 2630 acctctgagaagaagctaa 2648
 Db 2921 AATTAATTAAGAGGCCAAA 2939

RESULT 28
 PFAMSP1 PFAMSP1 1203 bp DNA linear INV 26-APR-1993
 LOCUS

DEFINITION	Plasmodium falciparum merozoite surface protein 1 (MSP1) gene, 3' end.
ACCESSION	M64681.1 GI:160538
VERSION	M64681.1
KEYWORDS	merozoite surface antigen; merozoite surface protein.
KEYWORDS SOURCE	Plasmodium falciparum (strain 79/94) asexual blood stage DNA.
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Hemosporidia; Plasmodium.
AUTHORS	1 (bases 1 to 1203)
TITLE	Blackman,M.J., Ling,I.T., Nicholls,S.C. and Holder,A.A. proteolytic processing of the Plasmodium falciparum merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains
JOURNAL	Unpublished (1991)
FEATURES	Location/Qualifiers
source	1..1203
gene	/organism="Plasmodium falciparum"
	/db_xref="taxon:5833"
	/dev_stage="asexual blood stage"
	1..1203
	/gene="MSP1"
CDS	1..1203
	/partial
	/gene="MSP1"
	/codon_start=1
	/product="merozoite surface protein 1"
	/protein_id="AAA29709.1"
	/db_xref="GI:160539"
	/translation="YIIVFGESREDDYDIDGQVVGGAVPYVIDNLTKIENEYEV
	ILKPLACIVYSLKQLENNWMTFNVNVDKILNSRFKRFKRWLESDLIPKIDLS
	SNYVVRDKPLNKKRDKPLSSYNYIKDSIDTDINEANDLVGYKKILSEYKSDLS
	IKKYINDROGENEKYILPELNNIETLYKTVNDRIDLFLHLEKAVLNTYEEKSNEVRI
	KELANLKTIOKRLADFFKKNFNVGIAIDSTIYNHNHLTKFLSTGVFEINAKTVLSN
	LIDNGLOGLMNTSOHCYVKQCPQNSGCFRILDEERECKCLANTKQEDKCEVNPNT
	CENNNGGCDADAKKCTHEEDSGSNKKITTECRKPPSYPLFDSIFCSSNFIISPLIL
	MLILXPSI"
sig_peptide	1..72
	/partial
	/gene="MSP1"
	73..858
mat_peptide	/gene="MSP1"
	/note="42 kDa fragment"
	/evidence="experimental"
	/product="merozoite surface protein 1"
	859..1200
mat_peptide	/gene="MSP1"
	/note="19 kDa fragment"
	/evidence="experimental"
	/product="merozoite surface protein 1"
BASE COUNT	496 a 143 c 171 g 393 t
ORIGIN	
Query Match	14.0% Score 691; DB 3; Length 1203;
Best Local Similarity	73.4% Pred. No. 1.3e-141;
Matches	883; Conservative 0; Mismatches 320; Indels 0; Gaps 0;
QY	3727 gtcacatgttgctacatcttcgcgagagagagagagactacagacctcgccaggtg 3786
DB	1 GTATCATAGTAGTACCTATATTGGAGAAATCCGAAAGAGATTATGATGATTTAGCAACAGTA 60
QY	3787 gtacccgggtgagggctgtgcaactcttcctgcgagatgattgataaatctgtccaaatcgagac 3846
DB	61 GTACACGGAGAAACACTAATCCTTCCTCGTAATTTGATTAACATACCTTTTAAAAATTTAAAAAT 120
QY	3847 gaatacgaagtgtctcatctgaacacctcgcgagcgctgctatcaggtctctcaagaacag 3906
DB	121 GAATATGAGAGTTTAAATTTAAACCTTTAGCAGCGGTTTATAGAAGTTTAAAAAACA 180
QY	3907 ctggagaaataacgtagtcacctcaatgtcaacgtgaaagacattctgacaagccgcttt 3966
DB	181 TTAGAAAAATTAACGTTATGACATTTAATGTAAATGTTAAAGATATTTTAAATTCAACGATTT 240

QY	3967	aataagagagaanaattccaagaagctcttgagagagcacttgattccctataagaacccg	4026
Db	241	AAATAAACCGTGAATAATTTCAAATAAGTTTAAATGATATCAATGATTAATGAAAGTTTAA	300
QY	4027	accctccctaactacagctgtgtccaaggagcccatcaagctccctcaataagaagaagagat	4086
Db	301	ACATCAAGTAATTAATGTTGTCAAAAGATCCATTAATTAATTTCTTAATTAATAAGAAAAGAGAT	360
QY	4087	aaattctgtcagtttacactatataccaagagctccatcgacacgcgataccaatttcgt	4146
Db	361	AAATCTTAAAGCAGTTATTAATTAATTAATAAGATTCATTAATGAAATGGAATTAATTTTGA	420
QY	4147	aatgatgtcgtgggtatttacagaagctccgagcgaanaatacaagctcacttgactc	4206
Db	421	AATGATGTTCTTGATATTAATTAATAATTAATTCGCAAAAATTAATAATTCAGATTACATTCA	480
QY	4207	attaaagaatatacaacagataagaagcgagaaataatcctgccttcctgaat	4266
Db	481	ATTAAAAAATATATCAACACCAACCAAGGTGAAATGAGAAATTAACCTTCCCTTTTAAAC	540
QY	4267	aacatcogaacccctgtacagacatgtgaacgcaaaatgactcctgttaattacccg	4326
Db	541	AATTAATGAACTTATATTAATAAACAGTTAATGATTAATTAATTAATTAATTAATTAATTA	600
QY	4327	gagcgcaagtcctcaactatacttacaagaagagcaatgtgtgaagttlaaaatcaagag	4386
Db	601	GAAAGCAAAAGTTCTAAATATATATATGAGAAATCAACGTAAGATTAAATTAATAAGAA	660
QY	4387	cgtgaactacctaaataacaatccaagaacaagctcgtgacagattccaagaataatacaattc	4446
Db	661	CTTAATTAATTAATTAATAAACCAATTAACAGCAAAATTTGCGATTTTAAATAATTAATTA	720
QY	4447	gtcggcaattgcagacctgtctacgcgaltlaacacaacaatcctcgtgccaagttctg	4506
Db	721	GTTGGATTGCGATTTATATCAACAGATTATTAACCAATTAATTAATTTGCAAAAGTTCC	780
QY	4507	lccactgtgcagtgtgtcgaanaacctcgcaaaacagtgctgtgaacaatcgtcgtgaacgc	4566
Db	781	AGTACAGGATGATGTTTGGAAAAATCTTGCTAAACCGTTTATCTTAATTAATTAATTA	840
QY	4567	aacgtgcagagcatcgtcgtacaatctcccaagacaacaatgtgtgaagaacagtgccccag	4626
Db	841	AACTTGCAAGGATATGTTAAACATTTTCACACACCAATGCGTTAAAAAACAAATGTCACACA	900
QY	4627	aatacgcgctgttccaagcatctgtgaacgagcgcggaagtgtaadgtgtctcctgaactac	4686
Db	901	AATTCGATGTTTTCAGACATTTAGATGAAGAAGAAATGTAATTAATTAATTAATTAATTA	960
QY	4687	aaacaagaagagataaagtgcgtgtgagaacccaacccactcgtcgaatgaataaacaatgac	4746
Db	961	AAACAAAGAAAGGATTAATAATGTTGAAAAATCCAAATCTTAATTTGTAACCAAAATTAATG	1020
QY	4747	gggtgtgacgcgcgatgtctaaatcgaccccgaggaagaacagcgtgtctaaagaaagaatc	4806
Db	1021	GGAT	1080
QY	4807	acatgcgagtgtaactaagcccgacactcctaactcctcgcagcggagattttgtccagc	4866
Db	1081	ACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1140
QY	4867	tctaatttccttgagatcctcctcctcgtcgtacatccatcgtatcgtgtgaagcattcac	4926
Db	1141	TCTAATCTTGTGAGATATCATTTTATTAATTAATCATATGTTAATTAATTAATTAATTAAT	1200
QY	4927	taa 4929	
Db	1201	TAA 1203	
RESULT 29			
LOCUS	AF325919	1131 bp	mRNA linear INV 13-FEB-2001
DEFINITION	Plasmodium falciparum merozoite surface protein 1 (MSP-1) mRNA.		

partial cds.
 AF325919
 AF325919.1 GI:12751398
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 malarial parasite *P. falciparum*.
 Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE
 AUTHORS
 TITLE
 1 (bases 1 to 1131)
 Li, X.R., Goel, V.K., Liu, S.C., Chishti, A.H. and Oh, S.S.
 42 kDa subfragment of MSP-1 gene of Plasmodium falciparum FCB-1 isolate
 JOURNAL
 REFERENCE
 2 (bases 1 to 1131)
 Li, X.R., Goel, V.K., Liu, S.C., Chishti, A.H. and Oh, S.S.
 Unpublished
 DIRECT SUBMISSION
 Submitted (04-DEC-2000) Section of Hematology/Oncology Research,
 St. Elizabeth's Medical Center, Tufts University School of
 Medicine, 736 Cambridge Street, Boston, MA 02135, USA
 location/Qualifiers
 1. 1131
 /organism="Plasmodium falciparum"
 /isolate="FCB-1"
 /db_xref="taxon:5833"
 /country="Colombia"
 <1. 1131
 /gene="MSP-1"
 <1. 1131
 /gene="MSP-1"
 /codon_start=1
 /product="merozoite surface protein 1"
 /protein_id="AAK07641.1"
 /db_xref="GI:12751398"
 /translation="AVTPSYIDNLSIKIENEYEVLYIKPLAGYRSIKOLENNWTF
 NVNKKDILNSREKRENFKNVLESLPIKDLSTNVVADPKFLKREKDFLSY
 NYTKSIDIDVDINADVLYKTLSEKYNKSDLSIKYINDKGENEKKYLPFLNLET
 LYKTVNDKIDLEFVILHLEAKVNTYEKSNVLEIKELNLTAKTIDKIDADKRNNEFV
 IADLSTVDYNNHNLTKFLSTGMFENAKVLEIKELNLTAKTIDKIDADKRNNEFV
 NSGCFHRLDERECKILNKGKGDCKVEPNPTCNNGCGCADAADKCTEDSGSNCK
 KINCECTKPSYFLPDGIFCSSNFIISLILMLILYFTI"
 BASE COUNT 467 a 136 c 154 g 374 t
 ORIGIN
 Query Match 13.3%; Score 655.8; DB 3; Length 1131;
 Best Local Similarity 73.7%; Pred. No. 7.4e-134;
 Matches 833; Conservative 0; Mismatches 297; Indels 0; Gaps 0;

D 361 GGATATTATAAATATTATTCGAAATAATATTAATACAGATTAGATTCATTAATAATAAT 420
 Q 4219 atcaagataagcaagcgagaaatgaaatatttgcctccctcgataatcaaac 4278
 D 421 ATCAACGCAAAAGGTGAAATGAGAAATACCTCCCTTTTAAACAAATTTGAAAC 480
 Q 4279 ctgtaacaagcagtaagcaaaatcgactctcgttaattcaacccgagcaagtc 4338
 D 481 TTATATTAACAGTTAATGATTAATGATTTATTTGTAATTCATTAGAACGAAAGTT 540
 Q 4339 ctcaactactacgaagaagcaatgtaagtaataaacaagcagtaactcc 4398
 D 541 CTAATATTACATATGAGAAATCAACGTAGAAATTAATAAGAACTTAATATTCTTA 600
 Q 4399 aaacaatccaagaagcagcagatttcaagaataataaatttcgcgcgaattgca 4458
 D 601 AAAACAATTCAGCAAAATGCGAGATTTTAAAAAATAACAATTCGTTGGAAATGCT 660
 Q 4459 gactgtctacgaattataacaacaatctcctgacaaagttctgtccactgcatg 4518
 D 661 GATTATCAACAGATTATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAAT 720
 Q 4519 gtgttcgaaacctcgccaaacagtgctgagcaatctgtcgaagcgaactgagggc 4578
 D 721 GTTTTGAATCTTGCTAAACCCGTTTATCTTAATTAATTAATTAATTAATTAATTAAT 780
 Q 4579 atgtctgaacatctcccgacaaatgctgtaagaagaagtgccccgaatagcgctgt 4638
 D 781 ATGTTAAACATTTTCACACCAACCAATGCTAAAAAACCAATTCACAAATTCGATGT 840
 Q 4639 ttcaagcactcgagcagcgagagtgcaagtgctcctgaaactcaacaagaagga 4698
 D 841 TTCGACATTTAGATGAGAAAGAGAAATGTAATGTTATTAATTAATTAATTAATTAAT 900
 Q 4699 gataagtcgtgagaaaccccaactcactgcaatgaaacaaatggcgggtgagccc 4758
 D 901 GATAAAGTGTTGAAATCCAAATCTGTAACCAAAATTAATGATGATGATGATGATGCA 960
 Q 4759 gatcttaatgacccggaggaagacagcgctcctgaagaaagaataacatgtagtgt 4818
 D 961 GATGCCAAATGTACCGAAGAGATTCAGATGCAACGCAAGAAATCAATCAATGATGT 1020
 Q 4819 actaagccgactcctcactcactcctcctcgaaggaatttttgcacagctcaattctctg 4878
 D 1021 ACTAAACCTGATTTCTTATTCACCTTTTCGATGATGATTTTTCGACATTCCTTA 1080
 Q 4879 ggcactcctcctcgtcgtatccatcgtcgtacgtcgtacagcttcaatca 4929
 D 1081 GGAATATCATCTTATTAATTAATCAATGATTAATTAATTAATTAATTAATTAATTAAT 1131

RESULT 30
 PFASURFPRO 1065 bp DNA linear INV 26-JUL-1993
 LOCUS
 DEFINITION
 Plasmodium falciparum merozoite surface protein (MSP-1) gene
 sequence.
 ACCESSION
 VERSION
 L20092.1 GI:309745
 KEYWORDS
 Plasmodium falciparum (strain Vietnam Oak Knoll) blood stage DNA.
 ORGANISM
 Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE
 1 (bases 1 to 1065)
 Louis-Willeman, V., Shi, Y., Collins, W. and Lal, A.
 Primary amino acid sequence of the carboxyl-terminal region of the
 merozoite surface protein (MSP-1) of Plasmodium falciparum Vietnam
 Oak Knoll (PVO) strain
 JOURNAL
 REFERENCE
 Unpublished (1993)
 location/Qualifiers
 1. 1065
 /organism="Plasmodium falciparum"
 /strain="Vietnam Oak Knoll"
 /db_xref="taxon:5833"

BASE COUNT 446 a /dev_stage="blood stage"
ORIGIN 125 c 150 g 344 t

Query Match 12.6% Score 623.6; DB 3; Length 1065;
Best Local Similarity 74.2%; Pred. No. 3e-127;
Matches 788; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

Oy 3799 gctgcactcctccgtgataataacatctgctccaaatcgagagacgaatacgaatg 3858
Db 1 GCAGTAACCTCCCTCGTATGTGATAACATCTTCTTAATAATGAAATATGAGGT 60
Oy 3859 cctctctgaaacctctgcaaggcgctctatagctctcgaagaacagctggaataac 3918
Db 61 TTATATTTAAACCTTACAGGTGTTTATGAGATTAAAAACAATTAGAAAAATAAC 120
Oy 3919 gtagagacctcaatgtaacgtaagacatctgaaacagcgcttaataagagaga 3978
Db 121 GTTATGACATTTAATGTTAATGTAAAGATATTTAAATTCACGATTTAATAACGTGA 180
Oy 3979 aattcaagaacgctctggaagcgactgattccctataagaacctgaccttaac 4038
Db 181 AATTTCAAAAATGTTTAAATCAGATTATTCATATAAGATTAAATCAATCAACTAAT 240
Oy 4039 taagttgcaaggaccacaaagtctcctaataaagaagaaggataaattctgctc 4098
Db 241 TATGTTGTCAAAGATCCATATTAATTTCTTAATAAGAAAAAGATTAATCTTAACG 300
Oy 4099 agttcaactatatacaagaccacacacgacatcaatcaatctgactgactc 4158
Db 301 AGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 360
Oy 4159 gggatltacaagatccctgagcgaanaatacaagctgacactgactc 4218
Db 361 GCATATTTAAATAATTTATTCGAAAAATTAATAATCGATTTAGATTTAAAAATAT 420
Oy 4219 atcaacgataagcgaaggcgagaaataatctgacctctcctgataatacgaacc 4278
Db 421 ATCAACGACAAACAGGTGAAAAATGAAAAATACCTTCTTTTAAACAATATGAGACC 480
Oy 4279 cgttacaagacgtgaagacacaacatcgacctctcgtatcaacctggaagccaagtc 4338
Db 481 TTATATTAATAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540
Oy 4339 ctaactactactaagagaagcaatgctggaagttaaaatcaaggagctgaactc 4398
Db 541 CTAATATTATACATATGAGAAATCAACGTAAGTAATAATAAGAACTTAATTA 600
Oy 4399 aaaaacaatccaaagacagctgcaagatctcaagaanaataaacaattcgctgaattgca 4458
Db 601 AAAACAAATTCAGCAAAATGCGAGATTTTAAAAAAAATAACAAATTCGTTGCAATTC 660
Oy 4459 gacctgctaacgaattatacaacaacatctcctgaccaaagttctgccaactggaatg 4518
Db 661 GATTTATCAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720
Oy 4519 gcttcgaanaacctcgccaanaacagtcgtaacaatctgctgaagcgaacctgcaaggc 4578
Db 721 GTTTTGAATAATCTTGCTAAACCGTTTATCTAATTTACTTGATGAAACTTGCAAGGT 780
Oy 4579 atgctgaacatctccagcaacaaatgctgaaagaacagtcgcccagaaatagcgctgt 4638
Db 781 ATGTTAAACATTTCAACACCAATGCGTAAAAAAAACAATGTCACAAATTTCTGGATGT 840
Oy 4639 ttcaagcatctgagagcggaagtgcaagtgctcctgaactacaacaagaaga 4698
Db 841 TTCAAGCATTTTGAATGAAGAGAGATGTAAATGTTTATTAATTTACAAACAGAGGT 900
Oy 4699 gataaagtgcgtggaagaccacaacctactcgaatgaaacaacatgcggtgctgacgc 4758
Db 901 GATTAATGTGTGAATAATCCAAATCTCTACTTGTAAAGAAATAATGTGTGTGTGATGCA 960

Oy 4759 gatgctaaatgcaaccggaagaacagcgctctaagcgaanaaataacatgcaagtgt 4818
Db 961 GATGCCAAATGTACCGAAGAGATTCAGGTACCAACGAAAGAAATCACATGTGAATGT 1020
Oy 4819 actaagcccgactcctatcactcctcgaaggaattttgc 4860
Db 1021 ACTAAACCTGATTCATTTCACACTTTTCGATGTGATTTTCTG 1062

Search completed: September 20, 2002, 02:49:10
Job time: 9339 sec